

HUMBOLDT-UNIVERSITÄT ZU BERLIN
INSTITUT FÜR BIBLIOTHEKS- UND INFORMATIONSWISSENSCHAFT



BERLINER HANDREICHUNGEN
ZUR BIBLIOTHEKS- UND
INFORMATIONSWISSENSCHAFT

HEFT 450

QUELLEN DIGITALISIEREN, DIGITALISATE ALS QUELLEN

ANFORDERUNGEN AN UND MANAGEMENT VON
RETRODIGITALISATEN IN GEDÄCHTNISEINRICHTUNGEN ALS
FORSCHUNGSDATEN DER GESCHICHTSWISSENSCHAFT

VON
JAN WIERZUCH

QUELLEN DIGITALISIEREN, DIGITALISATE ALS QUELLEN

ANFORDERUNGEN AN UND MANAGEMENT VON
RETRODIGITALISATEN IN GEDÄCHTNISEINRICHTUNGEN ALS
FORSCHUNGSDATEN DER GESCHICHTSWISSENSCHAFT

VON
JAN WIERZUCH

Berliner Handreichungen zur
Bibliotheks- und Informationswissenschaft

Begründet von Peter Zahn
Herausgegeben von
Vivien Petras
Humboldt-Universität zu Berlin

Heft 450

Wierzoch, Jan

Quellen digitalisieren, Digitalisate als Quellen : Anforderungen an und Management von Retrodigitalisaten in Gedächtniseinrichtungen als Forschungsdaten der Geschichtswissenschaft / von Jan Wierzoch. - Berlin : Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, 2020. - 200 S. : graph. Darst. - (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft ; 450)

ISSN 14 38-76 62

Abstract:

Forschungsdaten werden in der Regel als die Daten definiert, die die Grundlage der Forschung bilden, während des Forschungsprozesses entstehen oder dessen Ergebnis sind. In der Geschichtswissenschaft bilden Quellen auf unterschiedlichen Informationsträgern die Forschungsbasis, welche meist in Gedächtniseinrichtungen, also Museen, Bibliotheken oder Archiven, aufbewahrt werden. Im Zuge der digitalen Forschung kommt daher dem Digitalisieren dieser Quellen eine besondere Bedeutung zu, da dieses die Grundlage der digitalen Geschichtswissenschaft bis hin zu den Digital Humanities darstellt und die Quellen weiten Personenkreisen zugänglich gemacht werden können.

Diese Arbeit untersucht daher, wie historische Quellen digitalisiert, verwaltet und bereit gestellt werden müssen, um den Anforderungen der Geschichtswissenschaft an ihre Forschungsdaten gerecht zu werden. Dazu wurden zum einen qualitative Interviews mit für die Digitalisierung zuständigen Mitarbeitern aus Gedächtniseinrichtungen geführt, um hier Möglichkeiten und Vorgehensweisen bei Aspekten der Auswahl, Herstellung, Verwaltung und Nutzung der Digitalisate zu beleuchten. Hierbei wurden auch die digitalen Sammlungen der entsprechenden Einrichtungen berücksichtigt. Zum andern fanden qualitative Interviews mit wissenschaftlichen Nutzern statt, um deren Anforderungen an die Digitalisate, ihre Metadaten sowie die Verfügbarkeit und Nachnutzung zu betrachten.

Im Ergebnis zeigt sich hierbei, dass sich die Nutzer eine möglichst intuitive, übersichtliche Präsentation der Digitalisate wünschen, die jedoch alle vorhandenen Informationen bei Bedarf bereitstellen kann. Die Daten sollten dabei gut zugänglich und idealerweise verlinkt sein. Eine Beteiligung an organisatorischen Fragen der Digitalisierung oder technische Aspekte stehen nicht im Vordergrund, solange die Zugänglichkeit gegeben ist. Die Anforderungen entsprechen damit den bisherigen, analogen Erwartungen an Gedächtniseinrichtungen, Informationen zu sammeln, zu erschließen, zugänglich zu machen und zu sichern.

Diese Veröffentlichung geht zurück auf eine Masterarbeit im Masterstudiengang Information Science, M.A. an der Humboldt- Universität zu Berlin.

Eine Online-Version ist auf dem edoc Publikationsserver der Humboldt-Universität zu Berlin verfügbar.



Dieses Werk ist lizenziert unter einer [Creative Commons Namensnennung - Nicht kommerziell - Keine Bearbeitungen 4.0 International](https://creativecommons.org/licenses/by-nc-nd/4.0/) Lizenz.

Inhalt

1. EINLEITUNG.....	7
2. FORSCHUNGSDATEN IN DER GESCHICHTSWISSENSCHAFT.....	11
2.1. Forschungsdaten (allgemeine Definition).....	11
2.2. Forschungsdaten in der Geschichtswissenschaft.....	14
2.3. Digitale Entwicklungen zu den Digital Humanities.....	19
2.4. Forschungskreislauf der Geschichtswissenschaft.....	21
2.5. Bestehende Policies für Forschungsdaten in der Geschichtswissenschaft.....	22
3. DIGITALISATE ALS FORSCHUNGSDATEN.....	27
3.1. Datenmanagement der Digitalisate.....	29
3.2. Problematik Urheberrechte und Datenschutz.....	36
4. DIGITALISATE IN KULTURERBEEINRICHTUNGEN.....	43
5. UNTERSUCHUNG VON DIGITALISATEN IN KULTURERBEEINRICHTUNGEN.....	46
5.1. Vorstellung der Interviewpartner und Institutionen.....	46
5.2. Methodik der Interviews.....	48
5.3. Angebote und Möglichkeiten von Kulturerbeeinrichtungen.....	50
5.3.1. Digitale Sammlungen der untersuchten Institutionen.....	50
5.3.1.1. Methodik Webevaluation.....	50
5.3.1.2. Auswertung Webevaluation.....	51
5.3.2. Beschreibung der Fragen und des Aufbaus (Methodik) - Institutionen.....	54
5.3.3. Auswertung der Interviews - Institutionen.....	58
5.3.3.1. Auswahl (I-1.).....	58
5.3.3.2. Digitalisierung (I-2.).....	59
5.3.3.3. Dokumentation, Metadaten und Erschließungsstandards (I-3.).....	61
5.3.3.4. Veröffentlichung (I-4.).....	63
5.3.3.5. Interoperabilität und Schnittstellen, Portale (I-5.).....	65
5.3.3.6. Speicherung, Pflege und Archivierung (I-6.).....	67
5.3.3.7. Organisatorisches (I-7.).....	68
5.4. Welche Anforderungen haben Nutzer an Digitalisate als Forschungsdaten?.....	71
5.4.1. Beschreibung der Fragen und des Aufbaus (Methodik) - Nutzer.....	71
5.4.2. Auswertung der Interviews - Nutzer.....	73
5.4.2.1. System (N-1.).....	73
5.4.2.2. Anforderungen an Digitalisate (N-2.).....	74
5.4.2.3. Verfügbarkeit und Nachnutzung (N-3.).....	77
5.4.2.4. Vorstellungen und Wünsche (N-4.).....	79
6. ABGLEICH: WAS BIETEN EINRICHTUNGEN, WAS ERWARTEN NUTZER?.....	83
6.1. Anforderungen der Nutzer.....	83
6.2. Möglichkeiten der Institutionen.....	87
7. FAZIT: WIE SOLLTEN DIGITALISATE ALS FORSCHUNGSDATEN NUTZBAR GEMACHT WERDEN?.....	90
8. LITERATURVERZEICHNIS.....	92
8.1. Bibliographische Referenzen.....	92
8.2. Nicht-bibliographische Referenzen.....	99
9. ANLAGEN.....	102
9.1. Abkürzungsverzeichnis.....	102
9.2. Forschungskreislauf in der Geschichtswissenschaft.....	103
9.3. Forschungsdatenlebenszyklus in der Geschichtswissenschaft.....	104
9.4. Evaluationsfragen zur Evaluation der Websites.....	104

9.5. Evaluationsergebnisse der Websites.	105
9.5.1. Evaluation Stiftung Stadtmuseum Berlin, Sammlung Online.....	105
9.5.2. Evaluation Staatsbibliothek zu Berlin, Preußischer Kulturbesitz, Digitalisierte Sammlungen.	108
9.5.3. Evaluation Universitätsbibliothek der Humboldt-Universität zu Berlin, Digitale Sammlungen (digi-hub).....	110
9.6. Interviewfragen wie an die Interviewpartner gegeben.....	113
9.6.1. Interviewfragen der Institutionen.....	113
9.6.2. Interviewfragen der Nutzer.....	118
9.7. Interviewfragen mit Erläuterungen.....	121
9.7.1. Interviewfragen der Institutionen mit Erläuterungen.....	121
9.7.2. Interviewfragen der Nutzer mit Erläuterungen.....	129
9.8. Gespiegelte Fragen.....	134
9.9. Interviewtranskripte.....	135
9.9.1. Interview Stiftung Stadtmuseum Berlin.....	135
9.9.2. Interview Staatsbibliothek zu Berlin, Preußischer Kulturbesitz.	147
9.9.3. Interview Universitätsbibliothek der Humboldt-Universität zu Berlin.....	169
9.9.4. Interview Nutzer 1.	184
9.9.5. Interview Nutzer 2.	191
9.9.6. Interview Dr. Andreas Kohring.	192
9.10. Vorlage für Einverständniserklärungen der Interviewten.	200

1. Einleitung.

„Ich erinnere mich an einen Satz, als wir mit der Digitalisierung beginnen sollten. Da sagten die Kollegen, die das für eine Zumutung für unsere Bestände hielten: ‚Ach, und dann schmeißen wir unsere Original hinterher weg, ja?‘ Also ich denke, es ist für uns Bibliothekare kaum vorstellbar, dass wir unsere Bücher wegschmeißen. Die Originale ersetzen? [...] Wohl eher nicht.“ (Müller, SBB, I-3.2.)

Über Jahrhunderte wurden verschiedene kulturelle Objekte, Bücher und Schriftverkehr durch Gedächtniseinrichtungen, also Bibliotheken, Museen und Archive, gesammelt, erschlossen, für die Nutzung zugänglich gemacht und aufbewahrt. Sie sind damit nicht nur bedeutsam für die Erinnerungskultur von Nationen, Regionen, Orten oder Personen, sondern stellen mit ihren Objekten auch die Grundlage für verschiedene Forschungsdisziplinen der Geistes-, Sozial- und Kulturwissenschaften bereit, deren Ergebnisse sie wiederum aufbewahren und verfügbar machen (vgl. DBV 2018, S. 3 sowie S. 18 sowie S. 20).

Durch den digitalen Wandel seit den 1990er Jahren wurden nicht nur digitale Veröffentlichungen wichtiger, die Wissensrepräsentation findet seit dem zunehmend nur noch digital statt (vgl. Horstmann 2018, S. 108). Für die Gedächtnisinstitutionen bedeutet dies zum einen, dass sich aus der Wissenschaft nun nicht mehr nur die Ergebnispublikationen sammeln und nachnutzen lassen, sondern auch die Forschungsdaten, also die Daten, die während des Forschungsprozesses oder als sein Ergebnis in unterschiedlichen Formen entstehen (vgl. Kindling et al. 2013, S. 45). Zum anderen bieten sich durch die Digitalisierung nicht nur neue Möglichkeiten des wissenschaftlichen Arbeitens, durch Digitalisierungsprojekte werden auch analoge Objekte nun im Internet sichtbar gemacht, strukturiert und erschlossen und stehen so für die wissenschaftliche Nutzung zur Verfügung.

Auch wenn, wie anfangs dargestellt, Digitalisate die Originale innerhalb der Gedächtniseinrichtungen nicht ersetzen werden, könnten sie auf Grund ihrer leichteren und schnelleren Verfügbarkeit, der höheren Verbreitung und neuer Möglichkeiten der digitalen Verarbeitung für die Forschung wichtig und damit zu den (digitalen) Forschungsdaten der Geisteswissenschaften werden.

Im Unterschied zu anderen Forschungsdaten, die direkt aus der Forschung entstehen, werden Digitalisate durch die Kulturerbeeinrichtungen konkret für die Nutzer digitalisiert und in der Regel auch direkt durch die Einrichtung, nicht über Repositorien, angeboten. Durch diese direkte Kontrolle über die Digitalisate als Forschungsdaten ergeben sich jedoch auch verschiedene Fragen: Wie sollten Gedächtniseinrichtungen ihre Digitalisate aufbereiten und verwalten und was ist für sie überhaupt möglich? Könnten Digitalisate auf Dauer die Originale für die Forschung ersetzen? Und welche Anforderungen haben Geisteswissenschaftler an Digitalisate, die ihnen als Quellen, Forschungsdaten und

Hilfsmittel für ihre Forschung dienen sollen? Diese Arbeit wird sich mit der Geschichtswissenschaft als einer der geisteswissenschaftlichen Forschungsdisziplinen beschäftigen, so dass sich aus diesen Aspekten zusammengefasst die Frage ergibt, wie historische Quellen digitalisiert, verwaltet und bereit gestellt werden müssen, um den Anforderungen der Geschichtswissenschaft an ihre Forschungsdaten gerecht zu werden.

Zur Beantwortung der Frage wurden qualitative Interviews mit für die Digitalisierung zuständigen Mitarbeitern von Gedächtniseinrichtungen sowie mit wissenschaftlichen Nutzern geführt. Da diese Interviews sehr ausführlich auf die verschiedenen Aspekte der Auswahl, Herstellung, Verwaltung und Nutzung von Digitalisaten eingehen, wurden jeweils nur drei Interviews pro Gruppe geführt, so dass es sich hierbei eher um Fallbeispiele handelt, welche über keine allgemeine Repräsentativität verfügen können. Die Interviews wurden dabei mit Mitarbeitern der Stiftung Stadtmuseum Berlin, der Staatsbibliothek zu Berlin Preußischer Kulturbesitz und der Universitätsbibliothek der Humboldt-Universität zu Berlin geführt, die interviewten Historiker stammen aus der universitär-akademischen Geschichtswissenschaft, um sich auf den akademischen Forschungsdatenbegriff und die hiesige Forschungspraxis konzentrieren zu können. Auf die jeweiligen Interviewpartner wird an entsprechender Stelle genauer eingegangen.¹

Dabei gibt es ein leichtes Ungleichgewicht zwischen der Anzahl der Fragen in den beiden Gruppen, da sich die Anforderungen der Nutzer relativ leicht mit wenigen Fragen ermitteln lassen, während die Möglichkeiten und aktuelle Vorgehensweisen der Institutionen mittels verschiedener Fragen unter verschiedenen Aspekten betrachtet werden müssen, wie z.B. im Bereich des Forschungsdatenmanagements technische, rechtliche, organisatorische und inhaltlich-beschreibende Aspekte vertreten sind.

Da das Feld der Forschungsdaten sehr weit ist, soll am Anfang der Arbeit nur eine grobe Beschreibung der wichtigsten Aspekte und eine für die Arbeit genutzte Definition gegeben sowie die Forschungsdaten speziell für die Geschichtswissenschaft beschrieben werden, wobei verschiedene Aspekte wie Umgang und Austausch der Daten, die Rolle von Gedächtniseinrichtungen sowie die digitalen Entwicklungen innerhalb der Geschichtswissenschaft bis hin zu den Digital Humanities erläutert werden. Auch der typische Forschungskreislauf der Geschichtswissenschaft sowie bestehende Policies von Forschungsförderern für den Umgang mit Forschungsdaten in Bezug auf die Geschichtswissenschaft werden betrachtet.

Anschließend wird auf die Rolle von Digitalisaten als Forschungsdaten eingegangen, es werden Aspekte ihres Managements von Formaten über Metadaten, Qualität und Nachnutzung bis hin zur Langzeitarchivierung erläutert, hinzu kommen Probleme des Urheberrechts und des Datenschutzes bei der Digitalisierung, da „die Arbeit mit digitalen Inhalten [...] untrennbar mit einer erhöhten Sensibilität für Rechtsfragen verbunden“ (Steinhauer 2017, S. 347) ist. Im Anschluss werden die Besonderheiten von Digitalisaten als Forschungsdaten in Kulturerbeeinrichtungen betrachtet, die zuvor noch nicht behandelt wurden.

¹ Siehe Abschnitt 5.1. Vorstellung der Interviewpartner und Institutionen.

Der darauf folgende Teil behandelt die Interviews, wobei zunächst die Interviewpartner und die Methodik der Interviews vorgestellt werden. Anschließend wird zuerst auf die Angebote und Möglichkeiten der Kulturerbeeinrichtungen eingegangen, wobei sich ein kurzer Abschnitt mit Webevaluationen beschäftigt, da die jeweiligen digitalen Angebote der drei Einrichtungen für bestimmte Fragen untersucht wurden. Dann werden die Fragen beschrieben und anschließend die Antworten nach Fragengruppen sortiert ausgewertet. In dem gleichen Schema folgt danach eine Erläuterung und Auswertung der Fragen zu den Nutzeranforderungen.

Die Interviewauswertungen haben dabei eher beschreibenden Charakter, Rückschlüsse und Empfehlungen aus den zuvor ermittelten Möglichkeiten und Anforderungen werden durch einen Abgleich dieser im folgenden Abschnitt beschrieben, um darauf aufbauend abschließend die Frage nach Herstellung, Verwaltung und Bereitstellung von Digitalisaten als Forschungsdaten beantworten zu können.

Obwohl in der Literatur die Themen Forschungsdaten und Digitalisierung oft behandelt werden, ist zur Frage, ob Digitalisate Forschungsdaten darstellen oder wie sie in der Forschung genutzt werden, bisher relativ wenig vertreten. Ein ähnliches Thema findet sich zwar bei Irina Schubert (vgl. Schubert 2018), hier wird jedoch eher auf die Präsentation der Digitalisate, weniger auf ihre Nutzung und die Funktionalitäten oder Anreicherung, Nachnutzung und Verlinkung, alles für Forschungsdaten typische Aspekte, eingegangen. Dementsprechend spricht sie auch nur von digitalen Primärquellen und zieht keine Verbindung zum Bereich der Forschungsdaten. Zudem legt Schubert ihren Fokus auf Archive und Editionen.

Breit vertreten ist dagegen das Thema der Forschungsdaten und des Umgangs mit ihnen (vgl. Kaden 2018; vgl. Kindling et al. 2013; vgl. Simukovic et al. 2014), wobei sich auch zu den Teilbereichen der Forschungsdaten in den Geisteswissenschaften oder gar der Geschichtswissenschaft Werke zu verschiedenen Aspekten finden, wie z.B. dem Charakter der Daten in diesen Fachrichtungen (vgl. Andorfer 2015; vgl. Cremer et al. 2018; vgl. Sahle/Kronenwett 2013), ihrem Forschungsdatenlebenszyklus (vgl. Puhl et al. 2015) oder der Langzeitarchivierung in diesem Bereich (vgl. Pempe 2012).

Ebenso findet sich Literatur zur Digitalisierung der Geschichtswissenschaft (vgl. Hohls 2018), den digitalen Quellen in der Geschichtswissenschaft (vgl. Schaßan 2018) und den Möglichkeiten, die sich den Geisteswissenschaften durch die Digitalisierung ergeben (vgl. Crane et al. 2007; vgl. Given/Willson 2018). Digitalisate spielen vor allem bei Rechtsfragen eine Rolle (vgl. Diesterhöft 2014; vgl. Klimpel et al. 2017; vgl. Steinhauer 2017) oder im Bezug auf den Umgang mit digitalisierten Beständen in Gedächtniseinrichtungen (vgl. Stauffer 2016).

Obwohl die Themen Forschungsdaten und Digitalisierung international für alle Fachbereiche in Wissenschaft und Wissenschaftsverwaltung wichtig sind, „sind die Erinnerungsinstitutionen und Infrastrukturkomponenten überwiegend nach nationalen Gesichtspunkten organisiert, auch die Diskussionsnetzwerke orientieren sich an nationalen Communities oder phonetisch getrennten Gemeinschaften“ (Hohls 2018, S. 17). Auf Grund der Rolle von nationalen Infrastrukturen und Richtlinien für die

Forschungsförderung sowie des relativ eng begrenzten Rahmens der hier untersuchten Gedächtniseinrichtungen wird sich diese Arbeit daher zum größten Teil auf den deutschsprachigen Raum konzentrieren, auch wenn auf ausländische Aspekte hingewiesen wird oder sich einige Interviewfragen und -antworten durchaus übertragen lassen.

Auch der viel verwendete Begriff der Digitalisierung bedarf einer genaueren Definition, da darunter sowohl der Prozess der digitalen Ausrichtung einer Institution oder Gesellschaft verstanden werden kann, der Begriff aber auch die Umwandlung analoger Signale in digitale Formate beschreibt (vgl. Horstmann 2018, S. 95; vgl. Lang/Bohne-Lang 2019, S. 147). In der Arbeit wird mit dem Begriff der Digitalisierung in der Regel letzteres, also die Retrodigitalisierung, beschrieben, wobei die Begriffe Digitalisierung und Retrodigitalisierung hier gleichbedeutend verwendet werden.

Auch wenn die Zugänglichmachung von Objekten für die Allgemeinheit einen wichtigen Aspekt bei der Digitalisierung darstellt, soll der Fokus in dieser Arbeit auf den Forschungsdatenbegriff und damit den akademischen Kontext gelegt werden, was sich nicht nur in der Auswahl der interviewten Nutzer zeigt, sondern auch der besseren Vergleichbarkeit in Bereichen wie Datenmanagement, Nachnutzung und Erschließung dient. Auf Grund seiner Bedeutung gerade für Gedächtniseinrichtungen wird der Aspekt der allgemeinen Zugänglichmachung natürlich trotzdem behandelt.

Aus Gründen der Lesbarkeit wird in der Arbeit die männliche Form verwendet, welche aber selbstverständlich beide Geschlechter mit einschließt.

2. Forschungsdaten in der Geschichtswissenschaft.

2.1. Forschungsdaten (allgemeine Definition).

Daten sind unverzichtbare Ressourcen im Forschungsprozess. Das Konzept der Daten lässt sich dabei aus unterschiedlichen Perspektiven betrachten, wobei der umgangssprachliche Datenbegriff der Daten als elektronischen, binären Zeichenfolgen nur eine Perspektive darstellt. Auch rein analoge Ergebnisse naturwissenschaftlicher Forschung und objektive bis subjektive Beobachtungen stellen Daten dar (vgl. Voß 2013, S. 5f.). In den Naturwissenschaften kam es bereits früh zur Verarbeitung der hier anfallenden meist objektiven, faktenartigen Daten mit Hilfe von Computern. Diese Art der Daten prägte den damit entstehenden Begriff der Forschungsdaten, der Daten im Forschungsprozess (vgl. Baru 2007, S. 113; vgl. Voß 2013, S. 6). Zwar waren diese, meist umfangreichen, Datenreihen mittels Computern verarbeitet, die Publikation der Ergebnisse fand jedoch in der Regeln analog statt. Eine Publikation der Daten in analoger Form war uninteressant, da, neben gewaltigen Publikationskosten, eine nachträgliche Auswertung aller Daten kaum möglich gewesen wäre (vgl. Buckland 2017, S. 33-36).

Erst im Zuge der Digitalisierung seit der Mitte der 1990er Jahre sind wissenschaftliche Daten vor allem digital verbreitet. Damit wird der einfache Zugriff sowie das Verifizieren der Daten ermöglicht, so dass eine Nachnutzung dieser realistisch und daher der professionelle Umgang mit den Daten von großer Bedeutung ist. Wissenschaftler sind damit nicht nur Produzenten, sondern auch Konsumenten von Daten (vgl. Baumann 2014, S. 18; vgl. Martin 2013, S. 12; vgl. Schulze/Stockmann 2013, S. 34). Hinzu kommen die Möglichkeiten von digitalen Forschungsumgebungen oder Cyberinfrastrukturen, in denen die Daten von Forschern auch über große geographische Entfernungen hinweg kollaborativ verwaltet und genutzt werden können (vgl. Baru 2007, S. 113-116; vgl. Crane et al. 2007, S. 117). Auf Grund dieser Entwicklungen wird die Veröffentlichung von Forschungsdaten, neben der klassischen wissenschaftlichen Publikation von Ergebnissen, immer wichtiger (vgl. Kindling/Schirnbacher 2013, S. 130).

In dieser Arbeit soll die Forschungsdatendefinition von Maxi Kindling und Peter Schirnbacher zugrunde gelegt werden:

„Unter digitalen Forschungsdaten verstehen wir dabei alle digital vorliegenden Daten, die während des Forschungsprozesses entstehen oder ihr Ergebnis sind. Der Forschungsprozess umfasst dabei den gesamten Kreislauf von der Forschungsdatengenerierung, z. B. durch ein Experiment in den Naturwissenschaften, eine dokumentierte Beobachtung in einer Kulturwissenschaft oder eine empirische Studie in den Sozialwissenschaften, über die Bearbeitung und Analyse bis hin zur Publikation und Archivierung von Forschungsdaten. Digitale Forschungsdaten entstehen in allen Wissenschaftsdisziplinen und unter Anwendung verschiedener Methoden,

abhängig von der Forschungsfrage. Dies hat zur Folge, dass sie in unterschiedlichen Medientypen, Aggregationsstufen und Datenformaten auftreten. Um das Bereitstellen von Forschungsdaten und ihre Nachnutzung zu ermöglichen, sind Metadaten und eine Datendokumentation essentiell, die den Kontext der Forschungsdaten beschreiben sowie die Werkzeuge, mit denen sie erzeugt, gespeichert, bearbeitet und analysiert wurden.

Der gesamte Prozess, der die Allokation, die Generierung, die Bearbeitung und Anreicherung, die Archivierung und Veröffentlichung von digitalen Forschungsdaten selbst oder von einer entsprechenden klassischen Textproduktion unterstützt, wird inzwischen meist unter dem Begriff ‚Forschungsdatenmanagement‘ zusammengefasst“ (Kindling/Schirmbacher 2013, S. 130).

Das Forschungsdatenmanagement ist damit Teil des Informationsmanagements, der effizienten Verwaltung von verfügbaren Informationen, welches die Digitalisierung bzw. Planung und Beschaffung, die Verarbeitung, Erschließung, Bereitstellung und Archivierung von Informationen beinhaltet. Bezogen auf (digitale) Forschungsdaten betrifft dies damit die technischen Ressourcen, organisatorische und rechtliche Rahmenbedingungen, die Werkzeuge zur Verarbeitung sowie Publikation und Wissenschaftskultur (vgl. Kindling/Schirmbacher 2013, S. 131f.).

Dies bildet bereits einen einfachen Lebenszyklus von Forschungsdaten ab, welcher in der Regel aus den Schritten der Planung und Erstellung, Auswahl und Bewertung, der Übernahme der Daten in ein Archiv (Ingest), Speicherung, Durchführung von Erhaltungsmaßnahmen und schließlich Publikation bzw. Zugänglichmachung und Auffindbarkeit für Nutzer besteht, welche die Daten im Idealfall nachnutzen, also in ihre eigene Planung und Erstellung integrieren (vgl. Ludwig/Enke 2013, S. 15f.).

Wie in der Definition beschrieben, können Forschungsdaten in verschiedenen Aggregationsstufen auftreten. Am gängigsten ist dabei die Unterscheidung zwischen Roh-, Primär- und Sekundärdaten. Während die Rohdaten nicht aufbereitete Daten darstellen, welche z.B. als Ergebnisse einer naturwissenschaftlichen Messung erfasst worden sind, fand bei den Primärdaten bereits eine Aufbereitung, z.B. eine Bereinigung und Vereinheitlichung der Daten, statt, so dass diese im weiteren Forschungsprozess verarbeitet werden können. Sekundärdaten wurden dagegen bereits in anderen Kontexten verarbeitet und werden für die jeweilige Forschung nachgenutzt (vgl. Puhl et al. 2015, S. 9).

Da Forschungsdaten in allen wissenschaftlichen Bereichen und Methoden vorkommen, sind auch ihre Dateiformate sehr heterogen und können von gängigen Office-Anwendungen über statistische Datensätze, geographische Daten, Bild-, Audio- und Videoformate bis zu Spezialformaten jeweils verwendeter Tools reichen (vgl. Simukovic et al. 2014, S. 8).

Wurden Forschungsdaten auf Grund ihrer schlechten Publizierbarkeit lange Zeit nur als Arbeitsmittel aufgefasst, so führen die Möglichkeiten der elektronischen Publikation und Nachnutzung dazu, dass Fachzeitschriften zunehmend das Publizieren von

Forschungsdaten für veröffentlichte Artikel einfordern. Die Daten haben dabei nicht nur den Sinn, auch für andere Forschungen bereit zu stehen, sondern sollen auch den Arbeitsprozess der Forschung dokumentieren und Ergebnisse nachvollziehbar sowie reproduzierbar machen, um die Qualität von Forschung zu erhöhen (vgl. Andorfer 2015, S. 5; vgl. Ludwig/Enke 2013, S. 13; vgl. Simukovic et al. 2014, S. 10). Zudem führt die Publikation der Daten in der Regel auch zu einer Archivierung, womit die Daten dauerhaft gesichert werden können (vgl. Simukovic et al. 2014, S. 9).

Neben der Transparenz von Ergebnissen ist die Nachnutzung, wie bereits beschrieben, der weitere wichtige Aspekt der Forschungsdatenpublikation. Nachnutzung kann dabei sowohl die Nutzung fremder Forschungsdaten für die eigene Forschung als auch die Nutzung eigener Daten durch andere bezeichnen, als Data Integration auch über die Grenzen von Wissenschaftsdisziplinen hinweg (vgl. Baru 2007, S. 114; vgl. Simukovic et al. 2014, S. 10). Voraussetzung dazu ist immer eine klare und persistente Identifikation der Daten, um eine Zitierung in anderen Arbeiten möglich zu machen (vgl. Söllner 2017, S. 248). Besonders wichtig ist die Nachnutzung dabei in Disziplinen, deren Daten in der Form nicht reproduzierbar sind oder ein hoher finanzieller Aufwand hinter den Daten steht, wie es in Disziplinen der Fall ist, welche auf Forschungsgroßgeräte angewiesen sind (vgl. Baru 2007, S. 114-116). Zudem kann die Veröffentlichung der Daten zu einer Stärkung der Citizen Science führen, da die Allgemeinheit so Zugang zu den Daten erhält, was gerade im Bereich der Kultur- und Bildungseinrichtungen von Bedeutung ist (vgl. DBV 2018, S. 4).

Wichtig für die Nachnutzung ist neben der eigentlichen Publikation der Daten auch die Erlaubnis, diese nutzen zu dürfen, weshalb die Forschungsdatenpublikation oft mit der Idee des Open Access verknüpft ist. Dieses bezieht sich dabei nicht nur auf die Zugänglichkeit der Daten, sondern hat auch eine nutzungsrechtliche Komponente, so dass die Daten für alle (wissenschaftlichen) Zwecke uneingeschränkt genutzt werden dürfen (vgl. Horstmann 2018, S. 100; vgl. Steinhauer 2017, S. 350).

Zwar gibt es für die betroffenen Bereiche jeweils eigene Bezeichnungen, wie z.B. Open Data für Forschungsdaten, Open Source für Softwarecode etc., diese werden jedoch meist unter den Begriffen Open Access bzw. Open Science zusammengefasst (vgl. Horstmann 2018, S. 101).

Laut Umfragen sind viele Forscher dazu bereit, ihre Daten auch für andere Projekte zur Verfügung zu stellen, dennoch gibt es auch Vorbehalte gegenüber der freien Verfügbarmachung, welche sich vor allem auf der Sorge begründen, dass die eigenen Daten von anderen schneller bearbeitet und Ergebnisse publiziert werden könnten als durch den Datenurheber (vgl. Kindling et al. 2013, S. 56; vgl. Schulze/Stockmann 2013, S. 37).

Voraussetzung dafür, zugängliche und rechtlich verfügbare Daten in wissenschaftlichen Arbeiten zu nutzen, ist zudem die Qualität der Daten. Die Qualitätssicherung, welche die Qualität der Forschungsinfrastruktur, der Daten selber sowie der Aufbereitung der Daten berücksichtigt, stellt daher einen wichtigen Aspekt des Forschungsdatenmanagements

dar und kann nur durch die Wissenschaft selbst erfolgen (vgl. Kindling/Schirmbacher 2013, S. 133f.).

Da sich diese Arbeit auf Digitalisate aus Gedächtniseinrichtungen konzentriert, bei welchen auf Grund der Expertise der Mitarbeiter in den Institutionen von einer möglichst hohen oder zumindest angemessenen Qualität ausgegangen werden kann, soll hier nicht verstärkt auf die Qualitätskontrolle eingegangen, sondern nur die Grundlagen kurz erläutert werden.

Zwar ist „akademische Qualität [...] prinzipiell medienunabhängig“ (Nentwich/König 2010, S. 161), lange Zeit wurden digitale Publikationen jedoch als fragwürdig angesehen, da man ihnen eine intransparente Herkunft nachsagte. Publikationen sind im Internet schnell und günstig zu realisieren, zudem können Layout und Erscheinung von reputationsträchtigen Einrichtungen leicht nachgebaut werden. Gerade in der Anfangszeit war es zudem möglich, auch entsprechende URLs zu kaufen. Hinzu kam die Sorge, viele Online-Publikationen würden zu einer allgemeinen Qualitätsverschlechterung der Wissenschaft führen (vgl. Nentwich/König 2010, S. 144 sowie S. 155).

Durch Methoden der Integrität und Authentifizierung sowie neue Verfahren zur Qualitätssicherung wie Open Peer Review konnte diesen Sorgen entgegengewirkt werden. Ein weiterer Vorteil digitaler Publikationen liegt darin, dass Korrekturen nun auch nach der Veröffentlichung (ex post) möglich sind, solange diese Veränderungen dokumentiert werden (vgl. Nentwich/König 2010, S. 147f.).

2.2. Forschungsdaten in der Geschichtswissenschaft.

Viele Aspekte des Forschungsdatenkonzeptes sind aus den Naturwissenschaften entstanden, lassen sich aber dennoch auf die Geisteswissenschaften, besonders auf die Geschichtswissenschaft, übertragen, obwohl der Begriff der Forschungsdaten mit einer Betonung auf Überprüfbarkeit und Nachvollziehbarkeit selbst kaum eine Rolle spielt (vgl. Andorfer 2015, S. 11; vgl. Schöch 2013, S. 2).

Während sich die Naturwissenschaften eher um Forschungszentren konzentrieren, welche die Daten erstellen und aufbereiten (vgl. Horstmann 2018, S. 103), entstehen die Daten in der Geschichtswissenschaft durch ein Zusammenspiel aus Gedächtniseinrichtung (Daten verfügbar machen) und Forschung (Daten aufbereiten).

Auch der Begriff der Primär- und Sekundärdaten kann, anders als teilweise behauptet (vgl. Sahle/Kronenwett 2013, S. 76), übertragen, aber nicht gleichgesetzt werden. Werden in den Naturwissenschaften die Primärdaten oft mit quantitativen Daten gleichgesetzt (vgl. Kindling et al. 2013, S. 51), so sind die Untersuchungsgegenstände allein der Geschichtswissenschaft räumlich, zeitlich, thematisch und methodologisch zu weit gestreut, um sie auf qualitative oder quantitative Daten zu beschränken (vgl. Pempe 2012, S. 137). Wie im nächsten Abschnitt erläutert wird, kann jedes Objekt und damit alle Daten Untersuchungsgegenstand der Geschichtswissenschaft sein. Je nach Fragestellung können neu aus Quellen erschlossene Daten als Primärdaten untersucht werden oder als Sekundärdaten Aussagen stützen. Auch die Forschungsdaten anderer Wissenschaftler können als Sekundärdaten eigene Aussagen untermauern oder selbst untersucht und damit zu Primärdaten werden. Anders als in den Naturwissenschaften können Sekundärdaten damit auch wieder zu Primärdaten werden (vgl. Cremer et al. 2018, S.

145; vgl. Puhl et al. 2015, S. 9), für eine Zuordnung von Daten zu einer der beiden Gruppen ist in der Geschichtswissenschaft daher immer auch die zugehörige Fragestellung entscheidend.²

Grundsätzlich kann in der Geschichtswissenschaft jedes Objekt, analog wie digital, Überlieferung wie Artefakt, Untersuchungsgegenstand bzw. Quelle für die Forschung sein, es gibt also keine klar begrenzbaren Forschungsgegenstände³ (vgl. Kaden/Kleineberg 2016, S. 47f.; vgl. Reiche et al. 2014, S. 13). Zudem gilt das sogenannte Vetorecht der Quellen, wonach eine Quelle sämtliche angenommenen Vermutungen und Interpretationen widerlegen kann. Um den oben definierten Begriff der digitalen Forschungsdaten auf diese Objekte zu übertragen, sollte man diese jedoch eher als Forschungsobjekte beschreiben. Erst die daraus hervorgehenden digitalen Daten wie Digitalisate, Transkriptionen, Annotationen für Ideen, Thesen und Konzepten, strukturierte und kommentierte Bibliographien etc. wären damit die digitalen Forschungsdaten (vgl. Andorfer 2015, S. 11). Hierbei ist zu berücksichtigen, dass diese Forschungsdaten in der Regel erst durch die Bearbeitung der historischen Quelle im Hinblick auf eine Forschungsfrage entstehen und daher nicht mehr als objektive Rohdaten zu verstehen sind. Eine Ausnahme hierbei bilden Digitalisate in textueller oder grafischer Form: Da diese die ursprünglichen Quellen möglichst originalgetreu ohne Interpretationen wiedergeben sollen, könnten diese auch als digitale Rohdaten angesehen werden.

Die Geschichtswissenschaft ist eine hermeneutisch-interpretierende Forschungsform, Schreiben ist daher die Hauptkomponente geschichtswissenschaftlicher Arbeit, Monographien als Printpublikation überwiegend das Kommunikationsmittel von Forschungsergebnissen (vgl. Given/Willson 2018, S. 815; vgl. Hohls 2018, S. 22; vgl. Reiche et al. 2014, S. 8f.; vgl. Wissenschaftsrat 2012, S. 35-40). Die Quellen liegen in der Regel als Texte oder Bilder, teilweise auch als Video- oder Audio-Dokumente, selten als Statistiken oder Tabellen vor (vgl. Baumann 2014, S. 52; vgl. Kindling et al. 2013, S. 48f.). Die wichtigsten Arbeitsweisen neben dem Schreiben sind daher das Bibliographieren sowie das Exzerpieren/Notieren (vgl. Andorfer 2015, S. 8f.). Die Daten werden dabei durch das Sichten, Analysieren und Vergleichen der Quellen gewonnen. Dieser Prozess ist zwar auch analog möglich, heute finden diese Arbeitsschritte aber praktisch immer digital in Office-Tools oder Literaturverwaltungsprogrammen statt (vgl. Andorfer 2015, S. 8). Liegen die untersuchten Quellen noch in analoger Form vor, so erfolgt hier zugleich der Transformationsprozess ins Digitale. Digital vorliegende Quellen sind in der Regel Digitalisate, nur in einigen Fachbereichen wie der Zeitgeschichte können zunehmend auch Born Digitals als Quellen vorkommen. Der Großteil dieser Daten ist dabei qualitativ, nur in einigen Fachgebieten der Geschichtswissenschaft werden auch bzw. vorrangig quantitative Daten verarbeitet (vgl. Cremer et al. 2018, S. 145).

² In anderen Geisteswissenschaften kann dies anders sein, wenn sich diese nur auf spezielle Objekte als Forschungsgrundlage beziehen. So mag die Edition eines Manuskripts (Rohdaten) für die Literaturwissenschaft die Primärdaten darstellen, Anmerkungen anderer Forscher dazu die Sekundärdaten.

³ Für eine umfangreiche Auflistung siehe: DARIAH Wiki. Zusammenstellung geisteswissenschaftlicher Quellentypen, <https://wiki.de.dariah.eu/display/publicde/7.3+Zusammenstellung+geisteswissenschaftlicher+Quellentypen> (zuletzt abgerufen am 06.06.2019, 23:02).

Nach der Erstellung der Daten durch diese Arbeitsschritte ist auch hier eine Aufbereitung nötig, damit die Daten genutzt werden können, bspw. die Strukturierung von Text oder die Kommentierung von Bibliographien (vgl. Andorfer 2015, S. 11). Jedoch sind „Auswahl und Aufbereitung der Daten [...] in den Geisteswissenschaften an spezifische Fragestellungen, bestimmte disziplinabhängige Methoden und theoretische Perspektiven gebunden. Eine unmittelbare Nachnutzung [der Daten durch andere Forscher] ist daher nicht ohne Weiteres möglich“ (Cremer et al. 2018, S. 151).

Der Digitalisierung standen die Geisteswissenschaften lange Zeit skeptisch gegenüber, im Zuge der steigenden Qualität und Quantität digitaler Ressourcen wandelte sich diese Einstellung jedoch (vgl. Reiche et al. 2014, S. 4; vgl. Schubert 2018, S. 229). „Inzwischen sind digitale Verfahren auch in der geschichtswissenschaftlichen Forschung zu einer Selbstverständlichkeit geworden. Eine stetig wachsende Zahl von Quellen liegt inzwischen in digitaler Form vor, Informationen über Archiv- und Bibliotheksbestände sind deutlich leichter erreichbar als früher und die wissenschaftliche Kommunikation, das Exzerpieren, Ordnen, Dokumentieren, Schreiben als auch das Publizieren erfolgt zunehmend digital“ (Hohls 2018, S. 3). Dadurch haben auch die Geisteswissenschaften ähnlich den Naturwissenschaften inzwischen mehr Datenmengen zur Verfügung, als je manuell bearbeitet werden könnten (vgl. Crane et al. 2007, S. 117).

Zwar bieten sich durch die Digitalisierung Vorteile wie die Möglichkeit einer effizienteren Wissenschaft, vereinfachte sowie gleichzeitig erweiterte Recherchefunktionen oder die Möglichkeit der leichteren internationalen Vernetzung und Zusammenarbeit (vgl. Hohls 2018, S. 16f.), jedoch fand auch ein Paradigmenwechsel durch die Digitalisierung statt. Durch den leichten Zugriff auf idealerweise immer mehr Objekte scheint sich die Grundannahme der Geschichtswissenschaft, es sei tendenziell wenig überliefert, umgekehrt zu haben: Nun findet man zu praktisch allen Fragestellungen sowohl Quellen als auch bereits vorhandene Literatur. Für die Methode bedeutet dies, dass die Aufgabe nun nicht mehr im Finden von Material, sondern im Auswählen des richtigen Materials liegt (vgl. Haber 2012, S. 6; vgl. Rapp 2018, S. 264). Die klassischen Aufgaben der Forschung (Bibliographieren, Exzerpieren und Notieren) ändern sich dadurch jedoch nicht und digitale Quellen müssen genauso wie analoge einer Quellenkritik unterzogen werden, wobei Authentizität und Herkunft der Daten sowie ihre Entstehungsbedingungen untersucht werden müssen (vgl. Schaßan 2018, S. 1).

Auch in digitaler Form bleiben Text und Bild die zentralen Quellen der Geschichtswissenschaft, dennoch kommen neue Quellentypen durch die Digitalisierung hinzu. So können bspw. auch Metadaten genug Informationen enthalten, um untersucht zu werden, es kann zu einem verstärkten Einsatz von Normdaten für die Forschung kommen, historische Karten lassen sich mit modernen Geoinformationssystemen untersuchen oder Objekte werden als 3D-Modelle rekonstruiert, um daraus Erkenntnisse zu gewinnen (vgl. Given/Willson 2018, S. 810; vgl. Puhl et al. 2015, S. 11; vgl. Reiche et al. 2014, S. 14f.). Die Dateien können dabei im Aufbau z.B. als Text unstrukturiert, als XML-Datei semi-strukturiert oder als Datenbank voll-strukturiert vorliegen, hinzu kommen auditive oder visuelle Multimediaformate (vgl. Schöch 2013, S. 3; vgl. Pempe 2012, S. 143). Problematisch könnten Born Digitals aus dem Social Media Bereich als Quellen

sein, da deren Echtheit sich nur über informationstechnische Mittel wie Protokolle oder Codierungen überprüfen lässt. Zudem ist hierbei eine Dokumentation über die Metadaten besonders wichtig, bei einem Fehlen dieser ist deren Wert für die historische Forschung beschränkt (vgl. Fickers 2014, S. 26).

Interessant für eine weitere Verbreitung digitaler Methoden in den Geistes- und Kulturwissenschaften wäre auch die Etablierung von Virtuellen Forschungsumgebungen wie DARIAH-DE, welche digitale Forschungsmethoden verankern könnten und Infrastruktur, Software, Informationen sowie Beratung bieten (vgl. Fritze/Mikuteit 2012). An dieser Stelle kann jedoch nicht weiter auf diese Angebote eingegangen werden, da hier eine Unterstützung für die Forschung angeboten wird, die sich zu weit und umfangreich vom Thema der Arbeit entfernen würde.

Für zukünftige Historiker entsteht durch die Digitalisierung zudem das Problem, dass bereits im gegenwärtigen Alltag riesige Datenmengen produziert werden, welche in der zukünftigen Forschung ausgewertet werden müssen sowie ihre Authentizität überprüft und die historische Relevanz festgestellt werden muss (vgl. Fickers 2014, S. 26). Für diese Aufarbeitung können zwar zunehmend automatische Systeme eingesetzt werden, doch auch diese setzen voraus, dass die Daten gespeichert werden, was bei aktuellen Medien und Kommunikationsplattformen nicht unbedingt der Fall ist. Dieser Verlust an Quellen könnte in Zukunft starke Auswirkungen auf das kulturelle (digitale) Gedächtnis haben (vgl. Haber 2012, S. 6f. sowie S. 11).

Trotz der digitalen Entwicklungen scheint in der Geschichtswissenschaft noch immer ein geringes Bewusstsein für Forschungsdaten zu bestehen (vgl. Cremer et al. 2018, S. 143f.). Dafür gibt es verschiedene Gründe. Zum einen wurde der Begriff der Forschungsdaten lange Zeit nicht konkret für die Geschichtswissenschaft formuliert (vgl. Cremer et al. 2018, S. 153). Zum anderen bestand lange Zeit kein Bedarf an der Veröffentlichung von Forschungsdaten. Bei den qualitativ hermeneutischen Methoden der Geschichtswissenschaft steht das Verstehen und Interpretieren der Ergebnisse im Vordergrund, die Erkenntnisse sollen also direkt nachvollzogen und überprüft werden können. Eine Überprüfung mittels weiterer Daten als den Begründungen in der Arbeit scheint daher nicht nötig (vgl. Andorfer 2015, S. 11). Hinzu kommt, dass eine Publikation von Forschungsdaten nicht als wichtig für die Reputation von Wissenschaftlern angesehen wird. Für diese würde die Forschungsdatenpublikation damit nur einen Mehraufwand bedeuten, welcher sich nicht rentiert (vgl. Cremer et al. 2018, S. 156). Zudem werden die Forschungsdaten eher als Nebenprodukte der Forschung angesehen. War es früher nicht möglich, diese sehr langen Zwischenprodukte und Auswertungen in der gedruckten Publikation zu veröffentlichen, so ist durch die digitalen Entwicklungen eine Veröffentlichung der Forschungsdaten, z.B. in Form von Enhanced Publications, und damit eine direkte Verknüpfung mit dem publizierten Forschungsergebnis nun möglich (vgl. Cremer et al. 2018, S. 147f.). In der Praxis findet dies jedoch selten statt, die vorhandenen Repositorien dienen eher der Vernetzung und dem kollaborativen Arbeiten als der Dokumentation des Forschungsprozesses (vgl. Andorfer 2015, S. 12f.; vgl. Cremer et al. 2018, S. 144). Problematisch ist hierbei jedoch, dass die Repositorien, neben ihrer geringen Bekanntheit, nicht zentral erfasst sind, eine Publikation hier also kaum zur

Verbreitung der Daten in der Geschichtswissenschaft beiträgt (vgl. Cremer et al. 2018, S. 152).

Auch andere Aspekte des Umgangs mit Forschungsdaten sind eher zurückhaltend. Gebundene Monographien bleiben, trotz zunehmender Verbreitung von Open Access und neuen Qualitätssicherungsmethoden, weiterhin der Goldstandard für eine gute Reputation (vgl. Haber 2012, S. 10) und die Forschungsdaten werden, wenn überhaupt, durch individuelle Lösungen gesichert (vgl. Pempe 2012, S. 140).

Der Umgang mit Forschungsdaten kann dabei von Forschungsinfrastrukturen unterstützt werden. Der Begriff der Forschungsinfrastrukturen wurde auch aus den Naturwissenschaften übernommen und beschreibt hier vor allem Labore und Großgeräte. In den Geisteswissenschaften werden darunter Bibliotheken, Museen und Archive, also Gedächtniseinrichtungen, verstanden (vgl. Rapp 2018, S. 252). Diese sind die erste Anlaufstelle für physische wie digitale Informationen und die „Forschungsdatenlieferanten vieler Geisteswissenschaften, da für diese Disziplinen häufig die [...] vorhandenen Bestände als Forschungsgrundlage genutzt werden“ (Kaden/Kleineberg 2016, S. 46).

Dabei wird von den Gedächtniseinrichtungen nicht nur eine einfache Zugänglichkeit und Benutzung erwartet (vgl. Münster et al. 2018, S. 372), sondern auch, dass diese das kulturelle Erbe digitalisieren, strukturieren und erschlossen zur Verfügung stellen. Dies ist besonders wichtig, da die Digitalisierung nicht nur ein automatischer Datengenerierungsprozess ist, sondern domänenspezifische Anforderungen erfüllt werden müssen (vgl. Cremer et al. 2018, S. 147; vgl. Kaden/Kleineberg 2016, S. 49).

Wichtig für domänenspezifische Anforderungen ist auch die weitere Arbeitsweise der Forscher. So können durch digitale Technologien und die Unterstützung von Forschungsinfrastrukturen zwar kooperative bis kollaborative Arbeitsweisen unterstützt werden (vgl. Baru 2007, S. 115; vgl. Horstmann 2018, S. 96), doch in der Praxis hängen diese Möglichkeiten stark von der jeweiligen Disziplin ab. Gerade die Geisteswissenschaften arbeiten dabei oftmals zu lokal begrenzt, so dass ein Datenaustausch über weite Strecken, wie er durch das Internet ermöglicht wird, oftmals nicht sinnvoll ist, da sich die Schwerpunkte der Arbeit und damit die Anforderungen an die Daten unterscheiden. Zudem arbeiten die Geisteswissenschaften, und besonders die Geschichte, oft mit kulturellem Erbe, welches der lokalen Sprache zugeordnet ist. Ein internationaler Austausch oder gar internationale Kooperationen können jedoch nur über eine gemeinsame Sprache stattfinden (vgl. Baru 2007, S. 115; vgl. Crane et al. 2007, S. 119; vgl. Hohls 2018, S. 17).

Zwar kommt auch in den Geisteswissenschaften, gerade im Bereich der Digital Humanities, durch den Einsatz von neuen Technologien und Tools sowie großen Datensätzen zunehmend kollaboratives Arbeiten auf (vgl. Given/Willson 2018, S. 816), doch insgesamt sind kollaborative Arbeiten bei Historikern nicht verbreitet. Dies liegt an mehreren Gründen: Zum einen ist wie beschrieben weiterhin die Monographie als „Ergebnis eines solitären, individuell erarbeiteten Forschungsprozesses“ (Haber 2012, S. 10) der Goldstandard für die historische Wissenschaft. Zum anderen ist die Wissenschaftskultur hier weder auf digitale Publikationen noch auf die Nachnutzung von

Forschungsergebnissen auf Datenebene ausgerichtet, was sowohl an den meist für bestimmte Fragestellungen ausgewählten und aufbereiteten Daten liegt (vgl. Reiche et al. 2014, S. 27; vgl. Sahle/Kronenwett 2013, S. 82) sowie daran, dass „der Gedanke der freien Weiterverwertbarkeit durch Nutzung, Weiterverbreitung und Änderung ohne ausdrückliche Zustimmung des Autors [...] den Geisteswissenschaften weitgehend fremd“ (Söllner 2017, S. 249) ist.

2.3. Digitale Entwicklungen zu den Digital Humanities.

Bisherige digitale Veränderungen waren, wie im vorherigen Abschnitt geschildert, nur äußerlich, die Arbeitsweisen blieben gleich. Durch die nun mehr datengetriebene Geschichtswissenschaft könnten sich in Zukunft aber auch die Methoden verändern (vgl. Haber 2012, S. 11).

Erste digitale Methoden in der Geschichtswissenschaft gab es dabei schon seit der Mitte des 20. Jahrhunderts, angefangen durch Roberto Busa, der mit seinem Index Thomisticus eine corpus- und computerlinguistische Aufbereitung des Werkes Thomas von Aquin schuf und dabei unter anderem eine computergestützte Lemmatisierung des Werkes entwickelte (vgl. Baumann 2014, S. 11; vgl. Hohls 2018, S. 14; vgl. Rapp 2018, S. 253). Die computergestützte Textanalyse blieb auch in der Folgezeit wichtiges Forschungsfeld, hinzu kam die automatische Auswertung serieller historischer Daten. 1962 fand eine erste Konferenz zum Einsatz von Computern für die Geschichtswissenschaft und der elektronischen Verarbeitung von kulturellem Erbe statt (vgl. Haber 2012, S. 2). Die Arbeit mit Computern wurde dabei nicht als neue Methode, sondern als Arbeitserleichterung angesehen, Indices, Konkordanzen und das Erstellen von Editionen bildeten die Schwerpunkte in den 1960er bis 1980er Jahren (vgl. Haber 2012, S. 3; vgl. Rapp 2018, S. 253f.). In den 1990er Jahren wurde mit dem Aufkommen des World Wide Web die Informationsbeschaffung, -verarbeitung und Fachkommunikation wichtig für die Geschichtswissenschaft, ab 2000 fanden zudem erste eigene Publikationen von Forschungsdaten durch Historiker statt. Ab ca. 2005 kamen durch interaktive Dienste und Social Media neue Möglichkeiten hinzu, die die Zusammenarbeit und das Halten von Kontakten noch weiter erleichterten (vgl. Haber 2012, S. 3-5).

Die Idee zu digitalen Bibliotheken stammte auch bereits aus den 1960er Jahren, wurde jedoch konkret erst in den 1990er Jahren umgesetzt und verbreitete sich dann schnell (vgl. Saracevic 2004, S. 1). 1992 kam es zur Umsetzung des ersten Massendigitalisierungsprojektes durch IBM. Im Zuge des 500. Jubiläums der Columbus-Entdeckung wurden Materialien aus dem Archivo General de Indias in Sevilla digitalisiert. Dabei handelte es sich um einen einfachen Bildscan in Graustufen mit 100 dpi (vgl. Hohls 2018, S. 15). Hinzu kam die Volltextdigitalisierung durch TEI/XML-Kodierung (vgl. Rapp 2018, S. 254). Ab der Jahrtausendwende kam es zudem zu einer zunehmenden Verbreitung von originären E-Journals und E-Books (vgl. Lang/Bohne-Lang 2019, S. 147). Interessant ist in diesem Zusammenhang, dass bis Mitte der 2000er Jahre historischer Content den stärksten Zuwachs aller Inhalte des WWW aufwies (vgl. Hohls 2018, S. 7).

Inzwischen ist digitales Arbeiten in allen Geisteswissenschaften weit verbreitet, digitale Methoden finden, trotz der bereits frühen Ansätze, jedoch meistens noch eher für die

Sichtbarmachung von digitalisierten Daten als für die eigentliche Forschung Verwendung (vgl. Fickers 2014, S. 27f.).

Eine breite Anwendung digitaler Forschungsmethoden für geisteswissenschaftliche Themen findet dagegen in den Digital Humanities statt, welche eine Schnittstelle zwischen den geistes- und kulturwissenschaftlichen Disziplinen, der Informatik und der Informationswissenschaft bilden (vgl. Hohls 2018, S. 3; vgl. Reiche et al. 2014, S. 4). Eine klare Einordnung der Digital Humanities ist dabei jedoch nicht möglich, da sie geistes- und kulturwissenschaftliche Forschungsgegenstände mit Forschungsformen aus Naturwissenschaften, Informatik, Data Science und Informationswissenschaft verbinden⁴ (vgl. Given/Willson 2018, S. 808; vgl. Horstmann 2018, S. 97). Der Begriff der Digital Humanities tauchte erstmals 2004 auf, seit 2012 erfährt das Fach einen Schub, wobei der vorherige Charakter einer Hilfswissenschaft zu Gunsten der Verfachlichung zu einer eigenen Disziplin aufgegeben wurde (vgl. Hohls 2018, S. 4 sowie S. 6f. sowie S. 10).

Das Forschungsfeld der Digital Humanities ist dabei eng mit den verwendeten Methoden verbunden. Grundlage ist die elektronische Verfügbarkeit der Quellen, welche die Analyseverfahren wie Textanalysen und Text Mining, Netzwerkuntersuchungen und Geoinformationssysteme, aber auch Mustererkennungen erst ermöglicht (vgl. Fritze/Mikuteit 2012; vgl. Pempe 2012, S. 158f.; vgl. Reiche et al. 2014, S. 15 sowie S. 26; vgl. Voß 2013, S. 7). Durch die Verwendung von quantitativen Analysen steigt die Bedeutung von statistischen Auswertungen und Data Science im Vergleich zu den klassischen Geisteswissenschaften enorm. Dies hat jedoch auch zur Folge, dass die untersuchten Sammlungen groß genug sein müssen, da die Methoden bei zu kleinen Sammlungen ihre Robustheit und Genauigkeit einbüßen würden (vgl. Schöch 2013, S. 11).

In einem weiteren Kontext gehören auch Retrodigitalisierung, Archivierung, Datenmanagement, Verfügbarmachung und Anreicherung mit Metadaten und Normdaten zu den Aktivitäten der Digital Humanities, da sie deren Forschungsmethoden erst ermöglichen (vgl. Reiche et al. 2014, S. 6f. sowie S. 27f.).

Ein weiterer wichtiger Aspekt der Digital Humanities ist die Visualisierung ihrer Ergebnisse. Vertrauen die regulären Geisteswissenschaften eher auf Texte zur Wissensvermittlung, da Visualisierungen als zu ungenau angesehen werden (vgl. Given/Willson 2018, S. 813), werden sie in den Digital Humanities als Möglichkeit gesehen, Zusammenhänge aus großen Datenmengen greifbar darzustellen (vgl. Klein/Gold 2016; vgl. Reiche et al. 2014, S. 26). Im Zuge dieser Entwicklungen können auch neue Modelle für die Aufbereitung, Zugänglichmachung und Nutzung von Forschungsergebnissen ausprobiert werden (vgl. Ullmann 2016, S. 34).

Die Forschungsdaten sind sowohl Digitalisate, aber auch andere digitale Daten, wie z.B. statistische Auswertungen, Quellcode, Textkorpora, etc. Dies stellt insofern eine neue Facette in den Geisteswissenschaften dar, als dass die zielorientierte Generierung

⁴ Interessant ist auch, wie sich der Bereich der Data Science, welche sich bisher noch stark auf die STEM-Fächer und Informatik bezieht, im Zuge der Verbreitung der Digital Humanities entwickeln wird und welche Anwendungsfälle in den Geisteswissenschaften gefunden werden können. Eine weitere Beschäftigung hiermit würde sich jedoch zu sehr vom Thema der Arbeit entfernen.

maschinenauswertbarer Daten zu einem notwendigen Schritt in der Forschung wird (vgl. Cremer et al. 2018, S. 144; vgl. Steinhauer 2017, S. 350).

Interessant ist, dass ein sehr großer Teil der Forscher in den Digital Humanities die erhobenen Daten auch als nützlich für die Nachnutzung durch andere Wissenschaftler hält, was in den Geisteswissenschaften sonst eher nicht so stark verbreitet ist. Dementsprechend wäre auch ein großer Teil der Forscher bereit, eigene Daten verfügbar zu machen oder arbeitet durch aktives Datenteilen bereits kollaborativ (vgl. Baumann 2014, S. 56).

Die Verbreitung der Digital Humanities hat auch Auswirkungen auf die klassischen Geisteswissenschaften. So werden in der Geschichtswissenschaft nun neben den zunehmend digitalisierten klassischen Methoden auch neue digitale Methoden angewandt, wie z.B. die Programmierung eigener Skripte, statistische Auswertungen, Spracherkennungsverfahren oder das Erstellen von Ontologien für maschinenlesbare Verarbeitung (vgl. Given/Willson 2018, S. 811; vgl. Puhl et al. 2015, S. 23). Dennoch können solche Auswertungen bisher nur quantitative Ergebnisse ausgeben, für historisch relevante Aussagen muss weiterhin zusätzlich eine qualitative Auswertung oder Überprüfung stattfinden (vgl. Cremer et al. 2018, S. 145; vgl. Fickers 2014, S. 26).

2.4. Forschungskreislauf der Geschichtswissenschaft.

Der typische Forschungskreislauf in der Geschichtswissenschaft beginnt durch eine Fragestellung, welche durch das Lesen von Publikationen oder Quellen aufkommen kann. Diese Fragestellung wird, unter Zuhilfenahme entsprechender Literatur, in den wissenschaftlichen Kontext eingeordnet und ausformuliert, anschließend werden, meistens in den Infrastruktureinrichtungen der Geisteswissenschaften, den Gedächtniseinrichtungen, die entsprechenden Quellen und Literatur gesucht. Passende Quellen werden dann im Hinblick auf die gewählten Fragen analysiert und so gewonnene Daten unter Berücksichtigung von Sekundärliteratur ausgewertet. Anschließend kann mit den so gewonnenen Informationen die Forschungsfrage beantwortet und als Ergebnistext publiziert werden. Diese Publikation wird idealerweise von anderen Forschern gelesen und regt diese möglicherweise zu eigenen neuen Forschungsfragen an.⁵

Der Forschungskreislauf ähnelt dabei dem anderer Geisteswissenschaften (vgl. Given/Willson 2018, S. 808).

Der Forschungsdatenlebenszyklus der Geschichtswissenschaft unterscheidet sich dagegen etwas von denen anderer Disziplinen (vgl. Dierkes/Curdt 2018, S. 35), da die Quellen als Forschungsgrundlage bereits (analog oder digital) vorliegen. Eine konkrete Datenerstellung findet durch die Forscher daher nicht statt. Der Forschungsdatenlebenszyklus beginnt statt dessen direkt mit der Verarbeitung und Analyse der Quellen hin zu Exzerpten, Notizen oder ersten Theorien. Die Quellen als Forschungsgrundlage (Rohdaten) werden damit zu Primärdaten ausgewertet. Optional kann der Digitalisierungsprozess als erster Schritt angesehen werden, der in der Regel

⁵ Die Abbildung eines typischen Forschungskreislaufs der Geschichtswissenschaft findet sich im Anhang: 9.2. Forschungskreislauf in der Geschichtswissenschaft.

jedoch nicht durch die Forscher, sondern durch entsprechende Einrichtungen durchgeführt wird. Anschließend wird die Forschungsfrage mit Hilfe der Daten beantwortet und in einem Ergebnistext publiziert. Dieser Text enthält zwar Forschungsdaten, eine Veröffentlichung aller Daten ist in der Geschichtswissenschaft jedoch, mit Ausnahme von Editionen, unüblich. Findet doch eine Veröffentlichung statt, so werden die Daten hier archiviert, verfügbar gemacht und können in anderen Projekten nachgenutzt werden.⁶

2.5. Bestehende Policies für Forschungsdaten in der Geschichtswissenschaft.

Forschungsdaten kommen in allen Wissenschaftsdisziplinen vor, der Umgang mit ihnen ist aber keineswegs klar. Viele Forschungsförderer und Institutionen der Wissenschaftsorganisation besitzen daher inzwischen, teilweise auch disziplinäre, Richtlinien, Empfehlungen und Policies, die den Umgang mit den Forschungsdaten regeln, die durch ihre Förderung entstehen (vgl. Allianzinitiative 2018, S. 6). Die Policies umfassen dabei in der Regel Aspekte des Forschungsdatenmanagements, der Nachnutzung und Zugänglichmachung sowie der Aufbewahrung und Archivierung. Neben Forschungsförderern, die die Einhaltung der Richtlinien bei der Projektförderung fordern, kommen Anforderungen an die Forschungsdaten teilweise auch von Verlagen, welche die zugehörigen Texte publizieren (vgl. Beucke 2010, S. 21f.). Für Forscher stellen diese Richtlinien dabei keine Einschränkung dar, vielmehr sind verbindliche Regelungen oftmals gewünscht, da diese bei der Organisation helfen (vgl. Simukovic et al. 2014, S. 13).

Problematisch können solche Richtlinien jedoch sein, wenn sie sich nicht an den unterschiedlichen Fachkulturen orientieren. So sind z.B. viele Policies auf die Zeitschriftenartikel der Naturwissenschaften ausgerichtet, vernachlässigen dabei aber die Monographien der Geisteswissenschaften (vgl. Söllner 2017, S. 250f.).

Für diese Arbeit wurden daher einige Richtlinien von Forschungsförderern in Bezug auf die Geschichtswissenschaft betrachtet.

Die Allianz der deutschen Wissenschaftsorganisationen gibt die Empfehlung, freien Zugang zu Forschungsdaten, Publikationen und Werkzeugen gemäß der Berlin Open Access Erklärung herzustellen und diese nachnutzbar zu halten. Da die Daten und Metadaten großen Wert für die Wissenschaft haben, sollen sie möglichst lange zugänglich bleiben. Dazu sind (fachspezifische) Standards sowie eine Dokumentation nötig. Hinzu kommen Persistent Identifier (vgl. Allianzinitiative 2010, S. 2; vgl. Allianzinitiative 2017, S. 4 sowie S. 10; vgl. Allianzinitiative 2018, S. 5 sowie S. 11). Ein gutes Forschungsdatenmanagement, welches Erzeugung, Auswahl, Auswertung, Speicherung, Aufbereitung und Nachnutzung betrifft, sowie ein Datenmanagementplan werden empfohlen, da dieser Anforderungen bereits frühzeitig klärt und zudem beschreibt, welche Daten erhoben und wie sie gesichert werden sowie welche rechtlichen Regelungen gelten bzw. gelten, wenn sie veröffentlicht werden sollen (vgl. Allianzinitiative 2018, S. 4 sowie S. 7). Konkrete Empfehlungen für die Geschichtswissenschaft gibt es nicht. Die

⁶ Die Abbildung eines typischen Forschungsdatenlebenszyklus der Geschichtswissenschaft findet sich im Anhang: 9.3. Forschungsdatenlebenszyklus in der Geschichtswissenschaft.

Digitalisierung wird als das Erzeugen von Forschungsdaten aufgefasst (vgl. Allianzinitiative 2018, S. 4).

Das Bundesministerium für Bildung und Forschung (BMBF) beschreibt zwar neue Forschungs- und Kooperationsmethoden durch die Digitalisierung, hier fehlen aber noch Qualitätssicherungsmaßnahmen für Forschungsdaten, wie es sie z.B. bereits bei Fachzeitschriften gibt. Der Förderungsfokus liegt daher auf der Entwicklung und Erweiterung von Qualitätsstandards, standardisierten Metadaten, Interoperabilität und Nachnutzbarkeit. Die Standards sollen dabei disziplinspezifisch entwickelt werden, eigene disziplinspezifische Vorgaben gibt es dementsprechend nicht (vgl. BMBF 2018).

Als weitere Forderung wird jedoch eine Open Access Veröffentlichung genannt, auch möglich als Zweitveröffentlichung nach spätestens zwölf Monaten. Die Forschungsdaten sollen auch in Kultur und Wirtschaft verwendet werden können (vgl. BMBF 2018).

Auch die Deutsche Forschungsgemeinschaft (DFG) hat zwar keine konkreten disziplinären Anforderungen an das Forschungsdatenmanagement der Geschichtswissenschaft,⁷ sieht Digitalisate aber genau wie die Allianz der deutschen Wissenschaftsorganisationen auch als Forschungsdaten an und gibt hierzu sogar eine Reihe von Richtlinien und Empfehlungen heraus.

Zentral sind dabei die DFG-Praxisregeln zur Digitalisierung (vgl. DFG 2016), welche die technischen Voraussetzungen für DFG-geförderte Digitalisate festlegen. Auf Grund ihrer Bedeutung wird auf die Praxisregeln in Abschnitt 3.1. *Datenmanagement der Digitalisate* noch genauer eingegangen. Darüber hinaus wird von Digitalisaten als Forschungsdaten erwartet, dass sie mit hochwertigen Metadaten versehen sind, Normdaten verwendet werden und bereits vor Projektantrag eine Rechteklärung stattgefunden hat (vgl. DFG 2018, S. 5; vgl. DFG 2019b, S. 5f.). Erforderlich sind zudem die Zugänglichkeit auch über die Projektumgebung hinaus, z.B. in der Deutschen Digitalen Bibliothek oder dem Archivportal D, wenn rechtlich möglich als Open Access Angebot. Voraussetzung ist zudem die Zitierbarkeit der Digitalisate (vgl. DFG 2016, S. 41-43; vgl. DFG 2018, S. 5; vgl. DFG 2019b, S. 6f.).

Eine gewisse Steuerung der Digitalisierung findet auch durch die Ausschreibung der Projekte statt, in der spezielle Anforderungen und Schwerpunkte gestellt werden (vgl. DFG 2018, S. 2-5).

Darüber hinaus stellt die DFG weitere allgemeine Anforderungen an Forschungsdaten: So sollten die Daten unter dem Namen des erhebenden Forschers verfügbar sein und für die Nachnutzung mit Metadaten beschrieben sowie in einer sinnvollen Verarbeitungsstufe bereit stehen (vgl. DFG 2008, S. 3; vgl. DFG 2015, S. 1). Die Qualitätssicherung sollte nach den Kriterien der Fachcommunity erfolgen und bereits bei der Forschungsplanung sollte ein Konzept zur Rechteverwaltung und Nachnutzung erstellt werden, wobei Open Access zu bevorzugen ist. Die Daten sollen mindestens für zehn Jahre aufbewahrt bleiben (vgl. DFG 2008, S. 3; vgl. DFG 2015, S. 1f.).

⁷ Vgl. Deutsche Forschungsgemeinschaft. Umgang mit Forschungsdaten, https://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/nachnutzung_forschungsdaten/index.html#anker62194665 (abgerufen am 27.06.2019, 23:21).

Ebenso liegen von der Europäischen Kommission im Rahmen des Horizon2020 Programms Anforderungen an Forschungsdaten vor. So sollten diese den FAIR-Prinzipien (Findable, Accessible, Interoperable, Reusable) entsprechen und der Umgang mit den Daten durch einen Datenmanagementplan geregelt werden. Dieser beschreibt dabei den Umgang mit den Daten, welche Daten gesammelt und verarbeitet werden, welche Methoden und Standards Anwendung finden sowie Strategien zur Veröffentlichung und Sicherung (vgl. European Commission 2016, S. 3-5). Disziplinäre Anforderungen speziell für die Geschichtswissenschaft liegen nicht vor.

Außerdem gibt es von Seiten der Hochschulrektorenkonferenz (HRK) Empfehlungen zum Umgang mit Forschungsdaten an Hochschulen. Diese sind jedoch auch nur sehr allgemein gehalten und weisen darauf hin, dass Forschungsdaten, auch in den Geisteswissenschaften, zunehmend an Bedeutung gewinnen und für die Nachvollziehbarkeit von Ergebnissen wichtig sind. Deshalb sollten Leitlinien überuniversitär abgestimmt und Forschungsinfrastrukturen aufgebaut werden (vgl. HRK 2014, S. 2f.; vgl. Simukovic et al. 2014, S. 6f.).

Auch in anderen Ländern ist das Forschungsdatenmanagement ein wichtiges Thema. Zwar bietet das Institute for Museum and Library Services in den USA keine Hinweise zum fachspezifischen Umgang mit Forschungsdaten der Geschichtswissenschaft, aber durchaus zur Digitalisierung und Langzeitarchivierung von Objekten. Dies ist auch wichtig, da der größte Teil der digitalen Sammlungen in den USA von Bibliotheken (73%) und Museen (26%) bereitgestellt wird (vgl. IMLS 2019, S. Vf. sowie S. 37).

Im Gegensatz dazu sieht die Kommission Zukunft der Informationsinfrastruktur (KII) der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder Gedächtniseinrichtungen zwar als wichtige Informationsinfrastruktureinrichtungen im Bereich der Retrodigitalisierung, welche Wissenschaftler in allen Stufen des Forschungsprozesses unterstützen sollen, (vgl. KII 2011, S. 14f. sowie S. 21), als Produzentin von Forschungsdaten wird dagegen nur die Wissenschaft beschrieben (vgl. KII 2011, S. 29).

Ansonsten werden zwar Handlungsempfehlungen für die Arbeitsgebiete der KII (unter anderem auch Retrodigitalisierung und kulturelles Erbe, Open Access sowie Forschungsdaten) gegeben (vgl. Schulze/Stockmann 2013, S. 36), aber konkrete eigene Empfehlungen bleiben aus, es wird stets auf die fachlichen Regelungen verwiesen. Es finden sich allein einige fachübergreifende Rahmenbedingungen, auf die hingewiesen wird. So wird vor allem ein gutes Datenmanagement empfohlen, um die Daten verfügbar und nachnutzbar zu halten. Zudem sollten die Daten redundant gesichert und langzeitarchiviert werden und grundsätzlich als Open Access Dokumente zur Verfügung stehen. Als Digitalisierungsstandards wird auf die DFG-Richtlinien verwiesen (vgl. KII 2011, S. 14 sowie S. 38). Empfohlen werden zudem ein Datenmanagementplan und die Daten über Aggregatoren im Web zu verbreiten (vgl. KII 2011, S. 43f. sowie S. 144).

Der Rat für Informationsinfrastrukturen (RfII) als wissenschaftliches Beratungsgremium der Politik gibt ähnliche Empfehlungen, benennt im Gegensatz zur KII aber auch explizit Digitalisate aus Gedächtniseinrichtungen als Forschungsdaten (vgl. RfII 2016, S. 15; vgl. RfII 2017, S. 3).

Ansonsten werden auch nur allgemeine Empfehlungen gegeben, wie die Daten nach den FAIR-Prinzipien zu behandeln und ihre Verfügbarkeit, Überprüfbarkeit sowie Nachnutzung zu gewährleisten sind, wobei die Daten dazu redundant gespeichert werden sollten (vgl. RfII 2016, S. 38 sowie S. 45f.). Zudem wird auf die Vorteile einer offenen Lizenzierung und eines Datenmanagementplans als Teil der guten wissenschaftlichen Praxis hingewiesen (vgl. RfII 2016, S. 59 sowie S. 69). Neu im Vergleich zu den anderen Einrichtungen sind die expliziten Hinweise darauf, Daten auch nach Projektende weiter betreuen zu müssen sowie auf rechtliche Probleme im Internet, da hier deutsches Recht nicht unbedingt gilt (vgl. RfII 2016, S. 2 sowie S. 10).

Disziplinspezifische Empfehlungen werden nicht gegeben, es gibt nur die Hinweise darauf, dass sich bisheriges Forschungsdatenmanagement vor allem an aktiven, datenbasierten Vorreitercommunities orientiert, es jedoch wichtig wäre, sich zumindest innerhalb einer Disziplin auf gemeinsame Ziele und Standards zu einigen (vgl. RfII 2016, S. 52 sowie S. 60).

Der Wissenschaftsrat (WR) als politisches Beratergremium gibt sogar etwas strikere Empfehlungen ab. So sollten Projekte nur gefördert werden, wenn die langfristige Sicherung und Verfügbarkeit der Daten, welche mindestens nach den DFG-Richtlinien digitalisiert werden sollten, gesichert ist (vgl. WR 2012, S. 12). Hinzu kommen die Empfehlungen, einen Datenmanagementplan aufzustellen und publizierte Forschungsdaten mit einer DOI auszustatten (vgl. WR 2012, S. 14 sowie S. 57).

Digitalisate, speziell Bilddigitalisate, werden auch explizit als Forschungsdaten genannt (vgl. WR 2012, S. 18 sowie S. 38), es gibt sogar den Hinweis, dass diese, zusammen mit Born Digitals, an Bedeutung für die hermeneutisch-interpretierende Forschung gewonnen haben und auch einen Mehrwert gegenüber dem Original aufweisen können, z.B. durch Volltexterschließung, Bilderkennungsverfahren oder digitale Verlinkung (vgl. WR 2012, S. 36 sowie S. 51). Neben dem Hinweis auf die DFG-Richtlinien gibt es aber keine Anforderungen an die Digitalisate, schon gar nicht auf disziplinspezifischer Ebene.

Insgesamt zeigt sich hier, dass in den Empfehlungen und Richtlinien nicht speziell auf die Geschichtswissenschaft eingegangen wird. Öfters wird zwar auf fachliche Regelungen verwiesen, diese existieren für die Geschichtswissenschaft jedoch noch gar nicht und sind auf Grund der noch eher traditionellen Forschungspraxis auch im Alltag kaum zu finden. Statt dessen werden allgemeine Empfehlungen gegeben. Alle Einrichtungen heben dabei die hohe Bedeutung von Forschungsdaten hervor und fordern in der Regel ein klares Forschungsdatenmanagement, welches Auswahl, Erzeugung, Aufbereitung, Nachnutzung, Speicherung und Archivierung umfasst und oft auch die Verwendung von Persistent Identifiern und Normdaten beinhaltet. Auch die Verwendung von Open Access und (fachspezifischen) Standards für Qualität und Metadaten werden sehr oft genannt, ohne dabei konkrete Anforderungen zu stellen. Konkrete Anforderungen werden nur von

der DFG in ihren Praxisregeln zur Digitalisierung gegeben, auf die sich einige andere Einrichtungen beziehen. Digitalisate werden dabei in der Regel als Forschungsdaten anerkannt.

Allerdings scheinen sich viele der Richtlinien eher auf die Datenverwaltung durch Forscher zu beziehen, da sich einige Probleme wie die Verfügbarmachung, wenn möglich als Open Access, oder die langfristige Sicherung bei Institutionen und speziell Gedächtniseinrichtungen weniger stellen, da diese als öffentlich geförderte Einrichtungen in der Regel eh den Auftrag zur Erhaltung und Verfügbarmachung haben.

Zwar kommen bei den großen Forschungsförderern und Einrichtungen keine fachspezifischen Richtlinien für die Geschichtswissenschaft vor, eine Institution, die dennoch eine disziplinären Policy für die Geschichtswissenschaft bietet, ist das Historische Datenzentrum Sachsen-Anhalt am Institut für Geschichte der Martin-Luther-Universität Halle-Wittenberg,⁸ welches für die Bereitstellung, Aufbereitung und Auswertung historischer Quellen und Forschungsdaten zuständig ist. Für die Aufnahme von Daten ins Archiv werden hier die Anforderungen gestellt, dass die Übergabe der Daten in einem geeigneten Format, der Nutzungsrechte für die Daten sowie der Dokumentation und der Metadaten gewährleistet sind. Zudem werden eine dauerhafte Kontaktadresse, Regeln zum Umgang mit verwaisten Daten und (wenn durch den Lizenzvertrag nicht geregelt) Maßnahmen zur Überführung in langzeitgeeignete Formate benötigt. Für seine Nutzer bietet das Datenzentrum zudem Unterstützung bei der Metadatenaufbereitung sowie Beratung zu Analyse-Tools und Lizenzierungen.⁹

Insgesamt zeigt sich, dass es keine institutionsübergreifende disziplinäre Policy zum Umgang mit Forschungsdaten in der Geschichtswissenschaft gibt und auch nur wenige Forschungseinrichtungen eigene disziplinäre Policies haben. Historiker müssen sich daher meistens an allgemeine Forschungsdatenpolicies und Richtlinien der jeweiligen Einrichtungen halten. Ein großer Teil der Forscher tut dies auch und bewahrt z.B. Forschungsdaten gemäß den Empfehlungen zur guten wissenschaftlichen Praxis der DFG (vgl. DFG 2019a) für mehr als zehn Jahre auf (vgl. Kindling et al. 2013, S. 50).

⁸ Vgl. Historisches Datenzentrum Sachsen-Anhalt, <https://www.geschichte.uni-halle.de/struktur/hist-data/> (abgerufen am 01.07.2019, 23:39).

⁹ Vgl. Historisches Datenzentrum Sachsen-Anhalt. Workflow Datenmanagement, <https://www.geschichte.uni-halle.de/struktur/hist-data/datenmanagement/#anchor2919065> (abgerufen am 01.07.2019, 23:45).

3. Digitalisate als Forschungsdaten.

Da wie beschrieben jedes Objekt eine Quelle für die Geschichtswissenschaft sein kann, können auch alle Objekte relevant zum Digitalisieren sein. Der Begriff der Retrodigitalisierung wird in diesem Kontext dabei oft mit dem bloßen Scannen gleichgesetzt, ebenso wichtige Schritte sind daneben jedoch auch Erschließung, Präsentation und Zugänglichmachung (vgl. Ullmann 2016, S. 34), gerade, da ein Großteil der Digitalisate von Textquellen stammt und hier eine Aufbereitung zur maschinellen Verarbeitung Voraussetzung ist (vgl. Pempe 2012, S. 142).

Die Nutzung durch Historiker und andere Forscher stellt dabei aber nur einen der Gründe für die Digitalisierung dar. Zwei weitere, teilweise als noch bedeutsamer geschilderte Gründe (vgl. Fickers 2014, S. 25) sind die Sicherung der analogen Originale und die Verfügbarmachung der Objekte.

Da analoge Objekte verfallen, gehörten das Aufbewahren und Pflegen dieser schon immer zu den Aufgaben der Gedächtniseinrichtungen. Durch die Digitalisierung ist es nun möglich, sehr originalgetreue Kopien anzufertigen und nutzbar zu machen (vgl. DBV 2018, S. 18), wobei die Kosten dafür durch neue Techniken und günstigen Speicherplatz immer weiter sinken. Digitalisate bieten zudem verschiedene Nutzungsvorteile, wie die Möglichkeiten von Annotationen und Bemerkungen, Bearbeitungen und dem direkten Kopieren. Durch die Digitalisierung können aber auch Nachteile entstehen, wie das Veralten von Softwareformaten oder der Informationsverlust, der durch die Digitalisierung eintritt. Andere Formen wie die Microverfilmung sind zwar finanziell günstiger und länger haltbar, bieten aber nicht die gleichen Zugangsmöglichkeiten wie digitale Objekte (vgl. Fickers 2014, S. 25f.).

Gerade für die Fächer mit einer Bindung an unikale Überlieferungen stellt die Digitalisierung daher eine gute Möglichkeit dar, ehemals streng wissenschaftliche Archive der Allgemeinheit zu öffnen (vgl. Fickers 2014, S. 25; vgl. Rapp 2018, S. 253).

Dennoch bleibt die Nutzung von digitalen Objekten in der Wissenschaft ein wichtiges Ziel. Für eine sinnvolle Nutzung müssen die Digitalisate daher Anforderungen an die Heterogenität und Multimedialität der Daten erfüllen, gleichzeitig aber auch Historizität und Kontext der originalen Objekte darstellen (vgl. Cremer et al. 2018, S. 148). Diese qualitativen Eigenschaften (Integrität, Vollständigkeit, Kontext, Dokumentation) sind dabei für die Nutzung wichtiger als zusätzliche Tools, die mit den Digitalisaten angeboten werden (vgl. Schubert 2018, S. 230), da die Tools nicht unbedingt jede Anforderung erfüllen oder auch von anderen Stellen bezogen werden können, die Digitalisate jedoch die Grundlage der Arbeit bilden und in der Regel nur bei der anbietenden Institution zu finden sind. Daher sind auch Metadaten von entscheidender Bedeutung für den

wissenschaftlichen Wert von Digitalisaten (vgl. Fickers 2014, S. 26), auf diese wird daher gesondert im nächsten Abschnitt eingegangen.¹⁰

Für die Kontextsicherung können Digitalisate auch mit anderen Daten vernetzt (vgl. Kaden/Kleineberg 2016, S. 47) oder sogar zu digitalen Objekten verknüpft werden. Ein digitales Objekt kann dabei aus einem oder mehreren Einzelobjekten, auch in unterschiedlichen Dateiformaten, bestehen, auch ein Vermischen von Retrodigitalisaten und Born Digitals ist möglich (vgl. Puhl et al. 2015, S. 10f. sowie S. 143).

Bei der Erstellung und Publikation von Digitalisaten treten dabei neben geschichtswissenschaftlichen und archivatischen auch editorische Fragen auf, da „die Onlinepräsentation von digitalisiertem Archivgut [...] editionstechnisch gesehen vergleichbar mit einer Faksimiliausgabe“ (Schubert 2018, S. 223) ist. Auf genauere editorische Fragen kann an dieser Stelle nicht weiter eingegangen werden, es ist aber wichtig zu bedenken, dass sämtliche Funktionen, die in einer digitalen Edition eines Digitalisates vorhanden sind, auch gesichert werden müssen. Damit sind bei der Langzeitarchivierung nicht mehr nur die Daten, sondern auch die Nutzungsoberfläche betroffen, welche notfalls emuliert werden muss. Die sonst übliche Trennung zwischen Präsentations- und Archivierungsschicht ist damit nicht mehr möglich (vgl. Pempe 2012, S. 141).

Für die Forschung bieten Digitalisate dagegen viele Vorteile, da Objekte, deren Nutzung zuvor nur schwer und unter bestimmten Bedingungen möglich war, nun zumindest auf inhaltlicher Ebene frei zugänglich sind. Unter dem Einfluss der Digital Humanities nimmt dabei auch der Fokus der Geisteswissenschaften auf die Herstellung von Digitalisaten zu (vgl. Hohls 2018, S. 21), obwohl „davon auszugehen ist, dass nur ein Bruchteil der erfassten Digitalisate für geisteswissenschaftliche Forschungsfragen relevant ist“ (Pempe 2012, S. 148), wobei unklar ist, welche Digitalisate relevant werden könnten. Umgekehrt nimmt durch die großen Mengen an Digitalisaten die Bedeutung von automatischen Verarbeitungsmethoden und damit der Digital Humanities zu (vgl. Hohls 2018, S. 24).

Digitalisierte Quellen bilden daher zum einen eine Arbeitserleichterung für die Wissenschaft und beschleunigen z.B. durch ihre Zugänglichkeit die Forschung allgemein (vgl. Kaden/Kleineberg 2016, S. 47), zum anderen erhöht sich auch die Nutzungswahrscheinlichkeit einer Quelle, wenn diese bereits digitalisiert vorliegt. Müssen die Forscher dagegen erst die Digitalisierung anstoßen oder selber durchführen, so erhöht sich der Arbeitsaufwand und damit die Hürde (vgl. Given/Willson 2018, S. 812). Auch die Bekanntheit in der Fachcommunity ist für ein digitales Angebot weit wichtiger als eine allgemeine Sichtbarkeit im Internet (vgl. Czmiel et al. 2005, S. 3). Die Nutzeroberfläche sollte möglichst intuitiv sein und keine hohen technischen Anforderungen stellen (vgl. Münster et al. 2018, S. 372f.).¹¹

Auf die Möglichkeiten, die sich durch digitale Quellen für die Geschichtswissenschaft und die Digital Humanities ergeben, ist bereits eingegangen worden.¹² Im Allgemeinen ist aber festzuhalten, dass die Nutzung der Digitalisate sich je nach geisteswissenschaftlicher

¹⁰ Siehe Abschnitt 3.1. Datenmanagement der Digitalisate.

¹¹ Siehe hierzu auch Abschnitt 5.4.2. Auswertung der Interviews - Nutzer.

¹² Siehe Abschnitt 2.2. Forschungsdaten in der Geschichtswissenschaft sowie 2.3. Digitale Entwicklungen zu den Digital Humanities.

Fachrichtung unterscheidet. Während die Philologie vor allem auf eine korrekte Wiedergabe des Inhalts inklusive von Randbemerkungen an den richtigen Stellen Wert legt, ist für die Kunstgeschichte eine möglichst hochauflösende und farbechte Wiedergabe wichtig. Für die Geschichtswissenschaft ist dagegen der Kontext des Werkes ein zentrales Element, welches auch vom Digitalisat widergespiegelt werden muss (vgl. Reiche et al. 2014, S. 7f.; vgl. Simukovic et al. 2014, S. 37-40 sowie S. 48f.). Allgemein erwarten die Nutzer aber nicht nur Bilddigitalisate, sondern auch Strukturdaten und wenn möglich Volltext (vgl. Kaden/Kleineberg 2016, S. 50).

3.1. Datenmanagement der Digitalisate.

Das Datenmanagement von Digitalisaten beinhaltet eine ganze Reihe an Aufgaben, die sich größtenteils mit den Aufgaben des Forschungsdatenmanagements decken. Darunter werden alle Maßnahmen verstanden, um die Daten trotz Menge, Datenheterogenität und rechtlichen Hürden zu organisieren, nutzbar zu halten und zu pflegen (vgl. Baru 2007, 113; vgl. Kaden/Kleineberg 2016, S. 46; vgl. Ludwig/Enke 2013, S. 13).

Zu den immer wieder genannten Kernaufgaben gehören dabei das Verwalten der Dateiformate, die Beschreibung der erstellten Digitalisate mit Metadaten, Normdaten und einer Dokumentation, die Qualitätssicherung sowie Sicherstellen von Authentizität und Integrität, die Interoperabilität mit Hilfe von Standards, die Bereitstellung (auch über Schnittstellen) und Nachnutzbarkeit durch möglichst offene Lizenzen und das Identifizieren mit Hilfe von Persistent Identifiern sowie die Speicherung und Archivierung. Hinzu kommt die Bereitstellung einer technischen Infrastruktur, um das Datenmanagement überhaupt zu ermöglichen (vgl. Cremer et al. 2018, S. 149; vgl. Kaden/Kleineberg 2016, S. 52; vgl. Ludwig/Enke 2013, S. 33 sowie S. 39; vgl. Martin 2013, S. 12; vgl. Puhl et al. 2015, S. 29; vgl. Sahle/Kronenwett 2013, S. 82).

Ein möglichst standardisiertes Vorgehen beim Digitalisieren, sowohl innerhalb einer Einrichtung als auch zwischen mehreren Einrichtungen, sichert dabei eine möglichst hohe Qualität, da diese Standards erprobt und eingearbeitet werden können (vgl. Czmiel et al. 2005, S. 168). Eine klare Standardisierung und Strukturierung von Forschungsdaten erhöht zudem die Nutzerfreundlichkeit (vgl. Martin 2013, S. 20).

Technisch lassen sich dabei die Dateiformate, welche die Art der Daten und ihre technische Speicherung bestimmten, und Datenformate, welche die inhaltliche Struktur der Daten festlegen, unterscheiden (vgl. Schaßan 2018, S. 18). Dabei ist es wichtig, anerkannte Standards und Formate zu benutzen, welche sich je nach Anwendungsfall unterscheiden können, um die Daten möglichst lange nutzbar und speicherbar zu halten. Werden Daten bereits von Beginn an in standardisierten, empfohlenen Formaten gespeichert, so ist der Aufwand der Langzeitarchivierung weitaus geringer (vgl. Czmiel et al. 2005, S. 161; vgl. Puhl et al. 2015, S. 33).

Bei den Bildformaten unterscheidet man zwischen Rastergrafiken, welche das Bild aus einzelnen Pixeln zusammen setzen, und Vektorgrafiken, welche die Bilder aus geometrischen Formen berechnen. Vektorgrafiken bieten zwar einige Vorteile, wie z.B. die problemlose Skalierung von Bildern, können auf Grund der zugrunde liegenden

Formen jedoch nicht einfach aus einem Scan erzeugt werden und sind daher nicht für die Digitalisierung von Objekten geeignet (vgl. Schaßan 2018, S. 25).

Rastergrafiken setzen sich dagegen aus Pixeln zusammen und sind daher gut für aktuelle Scanverfahren geeignet, welche den Farbwert, zusammengesetzt aus den Rot-, Grün- und Blau-Werten, für jeden einzelnen Punkt ermitteln.¹³ Die Anzahl dieser Bildpunkte bestimmt dabei die Auflösung, gemessen als dpi (dots per inch). Die Pixelzahl sollte dabei so gewählt werden, dass noch das kleinste relevante Detail gut erkennbar ist, gängige Auflösungen sind 72 dpi für Web-Grafiken und 300 dpi für hochqualitative Digitalisate (vgl. DFG 2016, S. 6; vgl. Schubert 2018, S. 240).

Ein weiteres Qualitätsmerkmal für Rastergrafiken ist die Farbtiefe, welche angibt, wie viel Bit Speicher pro Farbkanal zur Verfügung stehen. Je mehr Bit zur Verfügung stehen, desto unterschiedlicher und genauer können die Farben wiedergegeben werden. Als Standard werden hier 24 Bit, also 8 Bit pro Farbkanal verwendet (vgl. Schaßan 2018, S. 16). Wichtig ist hierbei auch der Farbraum, welcher den Darstellungsbereich jedes Farbkanals definiert. Üblich sind hier die Farbräume AdobeRGB, ECI-RGBv2 der European Color Initiative sowie das etwas eingeschränkte sRGB (vgl. DFG 2016, S. 19), welche sich als Farbprofile in den Aufnahme- und Darstellungsgeräten einstellen lassen und so für eine gleichartige Farbwiedergabe sorgen.

Daneben ist das Format zentral für die Qualität von Digitalisaten, da hier eine mehr oder weniger starke Komprimierung stattfindet. Üblich für Rastergrafiken sind die Formate TIFF, JPEG und PNG, wenn primär Textquellen wie historische Drucke, Handschriften etc. digitalisiert werden auch PDF, wobei für eine wissenschaftliche Nachnutzung vor allem TIFF und JPEG empfohlen werden (vgl. DFG 2016, S. 10; vgl. Pempe 2012, S. 142; vgl. Schubert 2018, S. 240). Für hochqualitative, da verlustfreie Darstellungen wird dabei TIFF (Tagged Image File Format) verwendet, da dieses gut mit großen Farbräumen und umfangreichen Metadaten umgeht (vgl. Schaßan 2018, S. 17), für komprimierte Dateien das JPEG-Format (Joint Photographers Expert Group). Komplex und daher noch nicht weit verbreitet ist das JPEG2000-Format, welches eine verlustfreie Komprimierung bei hoher Qualität ermöglicht (vgl. Puhl et al. 2015, S. 33f.).¹⁴ Das PDF (Portable Document Format), auch in seiner Archivform PDF/A, ist eigentlich ein Vektorgraphikformat. Es wird aber vorrangig zur Darstellung von Texten verwendet und unterstützt auch eine Einbindung von Rastergraphiken. Der Vorteil liegt hierbei darin, dass die Darstellung auf allen Systemen gleich aussieht, so dass es vor allem für den Austausch oder die Veröffentlichung von Informationen genutzt wird. In PDF/A werden zudem alle darstellungsrelevanten Informationen gesichert, um eine korrekte Darstellung zu garantieren (vgl. Lang/Bohne-Lang 2019, S. 175f.).

Wie bereits beschrieben, hat die DFG mit ihren *Praxisregeln Digitalisierung* (vgl. DFG 2016) ein Regelwerk dazu herausgegeben, welche Anforderungen sie an die in den von ihr geförderten Projekten entstandenen Digitalisate stellt. Diese Praxisregeln haben sich zu einer Art Standard für die meisten digitalisierenden Institutionen entwickelt und sollen daher kurz genannt werden. Ziel der Praxisregeln ist es dabei, für Nachhaltigkeit bei den

¹³ Für genauere Angaben zu aktuellen Scanverfahren siehe: Lang/Bohne-Lang 2019, S. 156-169.

¹⁴ Für genauere Angaben zu den Formaten siehe: DFG 2016, S. 20f. sowie Lang/Bohne-Lang 2019, S. 174-176.

Daten, aber auch Vereinheitlichung zu sorgen (vgl. DFG 2016, S. 4f.). Als Ziel der Digitalisierung wird die „möglichst originalgetreue Wiedergabe des Objektes nach Maßgabe der wissenschaftlichen Erfordernisse“ (DFG 2016, S. 6) angegeben. Die zu verwendende Auflösung liegt daher bei mindestens 300dpi bei einer Farbtiefe von 24 Bit und einem AdobeRGB oder ECI-RGBv2-Farbraum. Als Format wird unkomprimiertes TIFF empfohlen, da die Formate TIFF-LZW oder JPEG2000 nicht lizenzfrei sind (vgl. DFG 2016, S. 6 sowie S. 19).

Im Vergleich zu Bilddigitalisaten sind Audio- und Videodigitalisate noch nicht so weit fortgeschritten, was auch daran liegt, dass Bibliotheken, welche lange Zeit die Vorreiter bei der Digitalisierung von Kulturgut waren, mehr mit optischen Medien arbeiten, während Audiosignale andere Techniken benötigen und höhere Anforderungen an die Pflege und Langzeitarchivierung stellen (vgl. Lang/Bohne-Lang 2019, S. 156; vgl. Pempe 2012, S. 149). Für Audiodateien haben sich dabei das WAV-Format (Waveform Audio File) und das MP3-Format (MPEG-2 Audiolayer III) durchgesetzt, für Videoformate können noch keine eindeutigen Empfehlungen gegeben werden (vgl. DFG 2016, S. 21). Langfristig ist jedoch auch mit einem höheren Aufkommen von Audio- und Video-Digitalisaten zu rechnen (vgl. Pempe 2012, S. 143).

Eine starke Verwendung gibt es dagegen bereits bei Textdaten. Für diese und die darin erfassten Strukturdaten gelten wiederum andere Anforderungen.

Zum einen gibt es hier eine Reihe an Dateiformaten, von welchen das reine Textformat TXT das simpelste darstellt. Hier existieren keine inhaltlichen Metadaten, die Informationen sind alle direkt im Textdokument gespeichert (vgl. Schaßan 2018, S. 5-7). Eine erweiterte Form stellen die Auszeichnungssprachen dar, welche den Text, seine Strukturen, Inhalte sowie Metadaten wiedergeben. Die Textteile werden dabei über Tags gekennzeichnet. Häufig kommen die Sprachen SGML (Standard Generalized Markup Language) bzw. das darauf aufbauende, leichter zu verwendende XML (Extensible Markup Language) oder im Internet HTML (Hypertext Markup Language) zum Einsatz. Vom Format her stellen diese auch Textformate dar und sind daher, ebenso wie Plain Text, gut für die Langzeitarchivierung geeignet (vgl. Schaßan 2018, S. 8-13). Hinzu kommen komplexe Formate wie gängige Office-Formate, EPUB, welches auf freien Standards wie XML basiert, und das bereits erwähnte PDF. PDF bietet zudem den Vorteil, dass sich hier Text und Bild überlagern lassen und so auch gescannte historische Schriften mit einem maschinenlesbaren Text hinterlegt sein können (vgl. Schaßan 2018, S. 13f.).

Zum anderen werden für die interne Struktur des Textes Auszeichnungsformate verwendet, welche meistens auf XML basieren. Der De-facto-Standard für Geisteswissenschaften ist dabei TEI (Text Encoding Initiative), welcher als Metasprache für die strukturierte Kodierung, Annotation, Darstellung, Interoperabilität und Archivierung von Dokumenten genutzt wird und dabei hard- und softwareunabhängig ist (vgl. Pempe 2012, S. 142f.; vgl. Rapp 2018, S. 255f.; vgl. Schaßan 2018, S. 19). Von der DFG wird TEI für die Erfassung der strukturellen Metadaten empfohlen (vgl. DFG 2016, S. 32). Eine andere Möglichkeit ist der ebenfalls auf XML basierende ALTO-Standard (Analyzed

Layout and Text Object), welcher besonders für die Darstellung von Layouts und für das Speichern von zugehörigen OCR-Texten genutzt wird. Dieser Standard wird von der Library of Congress entwickelt und von der DFG als OCR-Speicherformat empfohlen (vgl. DFG 2016, S. 37). Hinzu kommen spezielle Formate wie EAD (Encoded Archival Description), welches sich bspw. zur Abbildung von Hierarchien und Findmitteln in Archiven eignet (vgl. Schaßan 2018, S. 22) oder sehr simple Beschreibungsstandards wie Dublin Core (vgl. Sahle/Kronenwett 2013, S. 82).

Diese Textauszeichnungen betreffen dabei nicht nur das Layout, sondern auch die Semantik und Interpretation der Texte, so dass hier auch editorische Entscheidungen getroffen werden (vgl. Schaßan 2018, S. 18f.). Dies ist gerade bei der Darstellung der (maschinenlesbaren) Volltexte von Digitalisaten interessant, welche für eine wissenschaftliche Nachnutzung erforderlich sind. Erzeugt werden diese über verschiedene Verfahren wie das automatische OCR (Optical Character Recognition), manuelles Double Keying oder Mischformen daraus, gespeichert wie beschrieben in Formaten wie TEI oder ALTO (vgl. DFG 2016, S. 8 sowie S. 22; vgl. Pempe 2012, S. 142). Im Gegensatz zu PDF/A lassen sich diese Formate ohne zusätzliche Software zwar nur als einfache Texte darstellen, haben aber den Vorteil, dass sie Strukturinformationen, Fußnoten und Hierarchien in einem Textdokument speichern und mit entsprechender Software auch darstellen können (vgl. Puhl et al. 2015, S. 34).

Zentral für das Management von Digitalisaten sind ihre Metadaten, da sie ein einfaches Auffinden, Interpretieren und Nachnutzen der Daten ermöglichen (vgl. Baru 2007, S. 115). Stellt das Digitalisat eine Repräsentation eines Objektes dar, so sind die Metadaten die zugehörigen Beschreibungen und sollten dabei Informationen zum Objekt, den beteiligten Akteuren und dem Digitalisierungsvorgang bieten (vgl. Ludwig/Enke 2013, S. 63; vgl. Schöch 2013, S. 3).

Eine der Hauptaufgaben für Metadaten liegt dabei in der Verwaltung und Suche von Digitalisaten. Dies betrifft nicht nur Portale, um die Objekte hier klar differenzieren zu können (vgl. Baru 2007, S. 116), sondern auch die klassische Suche und vor allem Bildsuche, welche in vielen Systemen noch allein über eine Beschreibung der Bilder durch die Metadaten läuft (vgl. Münster et al. 2018, S. 368), da Content Based Information Retrieval durch aktuelle Entwicklungen zum maschinellen Lernen erst einen Aufschwung erfährt. Die Bedeutung der Metadaten, und auch die Bedeutung der Auswahl der richtigen Metadaten für ein Objekt (vgl. Münster et al. 2018, S. 375), steigt daher mit der Zunahme der Datenmenge in einem System (vgl. Czmiel et al. 2005, S. 27).

Die zweite große Aufgabe für Metadaten stellt die Dokumentation von Daten dar. Dies ist vor allem für die Nachnutzbarkeit der beschriebenen Daten wichtig, da Digitalisate als Forschungsdaten ohne Beschreibung ihres Kontextes kaum einen Wert für die Geschichtswissenschaft haben (vgl. Baumann 2014, S. 20; vgl. Fickers 2014, S. 26). Zum Kontext der Digitalisate gehören dabei zum einen die Kontextinformationen des originalen Objektes, um Interpretationen zur Entstehung, Provenienz und Überlieferung sowie Beziehungen zu anderen Objekten herstellen zu können (vgl. Schubert 2018, S. 222), zum anderen aber auch Kontextinformationen des Digitalisates, also eine Dokumentation zu dessen Herstellung und Verarbeitung. Dabei sollten Informationen zum

Digitalisierungsprojekt wie z.B. Auswahlkriterien für die Dokumente oder technische Entscheidungen wie Kameratyp und -einstellungen, Auflösung und Farbdarstellung, zum Status des Projektes oder zur Art der Transkription gegeben werden. Auch Informationen, wie vollständig das Dokument (Sichtbarkeit der Seitenränder, Umgang mit Beilagen) oder die zugehörige Sammlung digitalisiert ist, können enthalten sein (vgl. Schubert 2018, S. 227f.). Diese Informationen sorgen nicht nur für eine Verknüpfung von Digitalisat, Forschungsgegenstand und Kontext (vgl. Reiche et al. 2014, S. 25), sondern sie stärken auch das Vertrauen der Nutzer in die Dokumente sowie deren Authentizität für die wissenschaftliche Nutzung. „Je mehr Informationen zu einem Digitalisierungsprojekt zur Verfügung stehen, desto vertrauenswürdiger scheint die Quelle“ (Schubert 2018, S. 242 sowie vgl. S. 227). Zudem kann durch die Verarbeitung mit digitalen Methoden, z.B. aus dem Bereich der Digital Humanities, ein neuer Kontext entstehen (vgl. Ludwig/Enke 2013, S. 20), welcher so mit dokumentiert wird. Durch eine genaue Dokumentation des Digitalisierungs- und Verarbeitungsprozesses können damit zusammenhängende Informationen auch über das Projektende hinaus bewahrt werden (vgl. Pempe 2012, S. 156).

Auf Grund dieser Bedeutung von Metadaten ist ihre Erstellung ein wichtiger und notwendiger Schritt in der Digitalisierung (vgl. DFG 2016, S. 13), welcher in Beschreibungen jedoch oftmals hinter das Scannen oder Zugänglichmachen zurücktritt. Zwar ist das Erstellen der Metadaten, gerade wenn zusätzlich eine Erfassung der Strukturdaten oder von Volltexten erfolgen soll, mit hohem Aufwand verbunden (vgl. Hohls 2018, S. 12f.; vgl. Given/Willson 2018, S. 812), doch eine nachträgliche Beschreibung von Forschungsdaten mit deskriptiven Metadaten ist weit aufwändiger, die Kosten dafür liegen ca. dreißig mal höher als bei einer Erstellung der Metadaten gleich bei der Digitalisierung (vgl. Beucke 2010, S. 31). Das Problem fehlender Metadaten findet sich oft bei alten Forschungsdaten, welche noch ohne entsprechende Standards gespeichert wurden, und stellt hierbei neben veralteten Formaten das Haupthindernis bei der Nutzung dar (vgl. Pempe 2012, S. 151).¹⁵ Auf Grund von entsprechenden Standards und Anforderungen ist die Qualität von Metadaten in den letzten Jahren jedoch stark gestiegen (vgl. Münster et al. 2018, S. 380).

Bei der Qualitätssicherung von Retrodigitalisaten geht es vor allem darum, die Einhaltung technischer Standards und die Nähe des Digitalisats zum Original zu kontrollieren (vgl. Ullmann 2016, S. 41). Dabei werden meist von nationalen Wissenschaftsorganisationen Standards zu technischen Angaben wie Format, Auflösung, Farbtiefe etc. gegeben, die sich jedoch sehr ähneln und daher internationale Quasi-Standards bilden. So finden in den USA die Vorgaben der Federal Agencies Digital Guidelines Initiative (FADGI) Anwendung, in Deutschland die bereits oft erwähnten DFG-Praxisregeln zur Digitalisierung (vgl. Lang/Bohne-Lang 2019, S. 180). Zur Kontrolle der Ähnlichkeit zum Original und der Farbtreue wird mit farbkalibrierten Monitoren und zugehörigen Farbprofilen für Scanner gearbeitet, bei den Scannern werden regelmäßige Weißabgleiche zur Kalibrierung vorgenommen und die Kontrolle von Schärfe und Farbe

¹⁵ Siehe hierzu auch Abschnitt 5.3.3. Auswertung der Interviews - Institutionen.

der Digitalisate erfolgt zudem über spezielle Test Targets (vgl. Lang/Bohne-Lang 2019, S. 161).

Hinzu kommt eine Kontrolle der vergebenen Strukturdaten, der OCR-Erfassung und anderer Metadaten.¹⁶

Zur Qualitätskontrolle gehört dabei auch das Überprüfen von Authentizität, also der Echtheit des Digitalisates, und Integrität, der Unversehrtheit und Vollständigkeit.

Die Integrität von Daten kann z.B. bei der Übertragung beschädigt werden. Fehler lassen sich hier aber durch Hashwerte und Prüfsummen erkennen, so dass die Daten neu gesendet werden können (vgl. Schaßan 2018, S. 3).

Die Überprüfung der Authentizität als eine Grundvoraussetzung zur Nutzung der Digitalisate in der Geschichtswissenschaft (vgl. Schubert 2018, S. 229) gelingt vorrangig durch die oben beschriebene Dokumentation der Daten (vgl. Czmiel et al. 2005, S. 162), aber auch Aspekte wie der Ruf der veröffentlichenden Institution oder eines beteiligten Autors können entscheidend sein (vgl. Schaßan 2018, S. 2; vgl. Schubert 2018, S. 229).¹⁷ Dies ist besonders wichtig, da das Digitalisat im Internet nicht nur eine Kopie, sondern ein selbstständiges Objekt darstellt, welches sich durch Nachnutzung weiterentwickelt, so dass neue Verbindungen entstehen (vgl. DFG 2016, S. 41).

Ein weiterer wichtiger Aspekt des Datenmanagements von Digitalisaten ist die Interoperabilität der Formate, da diese die Suche in den Daten, den Datenaustausch, gerade bei Portalen, und das Arbeiten mit verfügbaren Metadaten ermöglicht (vgl. Baru 2007, S. 115). Zur Erhaltung der Interoperabilität gibt es dabei einige Empfehlungen. So sollten die Daten einfach, flexibel und nach den Standards der Community nutzbar sein. Wichtig sind auch die Fehlertoleranz sowie Unabhängigkeit von proprietärer Software und Verschlüsselungen, da diese nicht überall verfügbar sind und Probleme bei der Langzeitarchivierung bereiten können (vgl. Ludwig/Enke 2013, S. 22f.).

Anders als bei anderen Forschungsdaten ist die Nachnutzung bei Digitalisaten nicht nur eine weitere Nutzung nach der ursprünglichen ersten Verwendung bei der Erhebung, sondern in der Regel die erste wissenschaftliche Nutzung, da die Gedächtniseinrichtungen als Produzenten der Digitalisate diese nicht immer sofort selber nutzen. Dies hat zwar den Vorteil, dass die Digitalisate keinen bestimmten fachlichen Entstehungskontext haben, in dessen Fragenbereich sie sich allein anwenden ließen, auf Grund der hohen Kosten, die bei der Erstellung und Pflege der Digitalisate entstehen, kommt der Nachnutzung im Vergleich zu anderen Forschungsdaten aber eine erhöhte Rolle bei der Begründung dieser Kosten zu. Es wird erwartet, dass die Daten genutzt werden (vgl. . Ludwig/Enke 2013, S. 40).

Voraussetzung für eine möglichst breite Nutzung der Digitalisate ist dabei eine möglichst freie Rechtseinräumung, wie sie auch von der DFG erwartet wird (vgl. DFG 2016, S. 34). Diese setzt voraus, dass Einschränkungen des Open Access maximal 5% der Digitalisate eines Projektes ausmachen dürfen (vgl. DFG 2016, S. 10) und die Qualität dabei für eine

¹⁶ Siehe hierzu auch Abschnitt 5.3.3. Auswertung der Interviews - Institutionen.

¹⁷ Siehe hierzu auch Abschnitt 5.4.2. Auswertung der Interviews - Nutzer.

große Zahl der üblichen Forschungszwecke ausreichend sein muss (vgl. DFG 2016, S. 42).

Bei der Veröffentlichung der Digitalisate kann dabei zwischen Masterkopien in höchster Qualität und Gebrauchskopien für bestimmte Einsatzzwecke unterschieden werden (vgl. Czmiel et al. 2005, S. 159), unbedingt ist aber eine Kennzeichnung der Herkunft (im Sinne einer klassischen Autorenschaft von Forschungsdaten) sowie eine klare Adressierung erforderlich, um die Nutzung zu ermöglichen (vgl. DFG 2016, S. 7 sowie S. 41; vgl. Kindling et al. 2013, S. 59). Die Zugriffsmöglichkeiten sollten möglichst breit gestreut sein, z.B. über Angebote auf der eigenen Website, in Portalen, über Web-Suchmaschinen oder OAI-Schnittstellen, um möglichst gut auffindbar zu sein (vgl. Beucke 2010, S. 51; vgl. DFG 2016, S. 10 sowie S. 44).

Eine Adressierung sollte durch Persistent Identifier, möglichst auf Dateiebene und, wenn vorhanden, zusätzlich auf Objektebene, erfolgen (Puhl et al. 2015, S. 26), da dies zentral für eine mögliche wissenschaftliche Nachnutzung ist (vgl. Czmiel et al. 2005, S. 161).

Ein weiterer zentraler Teil des Managements von Forschungsdaten ist die Langzeitarchivierung und Erhaltung der Zugänglichkeit (vgl. Martin 2013, S. 15). Unter Langzeitarchivierung versteht man dabei die Bewahrung der digitalen Objekte auch über technologische oder soziokulturelle Veränderungen hinaus. Die Dauer der Langzeitarchivierung kann je nach Kontext eine definierte oder unbestimmte Zeitspanne betragen (vgl. Beucke 2010, S. 1; vgl. Oßwald et al. 2012, S. 16).

Die Forschungsdaten sollten dabei archiviert werden, um sie auch in Zukunft nachprüfen zu können. Dies gilt besonders für Daten, deren erneute Erhebung in der Form nicht möglich ist. Bei Digitalisaten geht es dagegen vor allem um den Erhalt der Investition an Zeit und Geld sowie darum, die originalen Objekte vor einem erneuten Digitalisierungsprozess zu bewahren (vgl. Beucke 2010, S. 2f.; vgl. Czmiel et al. 2005, S. 21). Die DFG fordert in ihren Regeln der guten wissenschaftlichen Praxis sowie ihren Leitlinien zum Umgang mit Forschungsdaten eine Archivierung der Daten für mindestens zehn Jahre (vgl. DFG 2015, S. 1; DFG 2019a, S. 22), gerade in der geisteswissenschaftlichen Praxis sind jedoch viel längere Phasen nötig, so dass die Daten auch darüber hinaus erhalten werden müssen (vgl. Cremer et al. 2018, S. 154).

Die Langzeitarchivierung von Forschungsdaten beinhaltet dabei nicht nur die Inhalte an sich, sondern auch den Kontext zu den Daten (welcher idealerweise in der Dokumentation beschrieben wird), um die Daten nutzbar zu halten (vgl. Baumann 2014, S. 22; vgl. Beucke 2010, S. 1). Um dabei das Veralten der Daten zu verhindern, ist eine Pflege der Daten (Data Curation) nötig, die diese auch aus fachwissenschaftlicher Perspektive nutzbar hält (vgl. Cremer et al. 2018, S. 154; vgl. Pempe 2012, S. 155). Veralten heißt hierbei nicht nur, dass die Dateiformate nicht mehr nutzbar sind, sondern auch, dass die Daten aktuellen Nutzererwartungen und Anforderungen nicht mehr entsprechen (vgl. Ludwig/Enke 2013, S. 36).

Dazu werden sowohl technische Betreuung sowie organisatorische Verwaltung der Daten benötigt (vgl. DFG 2016, S. 39). Auf technischer Seite lassen sich digitale Objekte auf physischer, logischer und konzeptueller Ebene, also als Bitstream, Dateiformat und auf der Interpretationsebene, betrachten. Für eine komplette Nachnutzbarkeit müssen diese

drei Ebenen erhalten bleiben (vgl. Ludwig/Enke 2013, S. 14). Die Erhaltung der konzeptuellen Ebene ist am einfachsten möglich, wenn alle betreffenden Daten in einem einheitlichen Datenmodell gespeichert sind, so dass die Strukturen und Verknüpfungen übernommen werden (vgl. Puhl et al. 2015, S. 14). Auf der Ebene der Dateiformate ist, gerade bei einer großen Anzahl von Daten im gleichen Format, die Emulation der benötigten Software möglich, um die Dateien weiter nutzen zu können. Bei wenigen Daten kann dagegen eine Formatmigration durchgeführt werden, welche die Inhalte in ein neues Dateiformat überträgt (vgl. Puhl et al. 2015, S. 32).¹⁸ Die Erhaltung des Bitstreams sichert den Erhalt der unversehrten Datei. Dies beinhaltet sowohl den Schutz vor Datenschäden, korrupten Dateien oder versehentlichem Löschen sowie vor Hardwareschäden (vgl. Czmiel et al. 2005, S. 34) und lässt sich z.B. über redundante Datenspeicherung sowie regelmäßige Medienmigration, gerade zum Ende der Lebenszeit eines Datenträgers, ermöglichen (vgl. Baumann 2014, S. 23; vgl. Czmiel et al. 2005, S. 179; vgl. Puhl et al. 2015, S. 32).¹⁹

Die Organisation der Langzeitarchivierung steht dabei vor dem Problem, dass viele unterschiedliche Dateiformate in der Pflege teuer sind. Es muss daher nicht nur geklärt werden, welche Daten in Zukunft von Wert sein können (vgl. Baru 2007, S. 113), sondern auch, welche Formate ideal für die Erhaltung sind (vgl. Beucke 2010, S. 30). Proprietäre oder verschlüsselte Formate eignen sich daher nicht für die Langzeitarchivierung (vgl. Beucke 2010, S. 10; vgl. Ludwig/Enke 2013, S. 36). Organisiert werden muss dabei nicht nur die Erhaltung, sondern auch die Möglichkeit des Zugangs und der Nutzung, da die Daten auch weiterhin genutzt werden sollen (vgl. DFG 2016, S. 39). Ein Modell zur Beschreibung eines solchen Archivs ist das Open Archival Information System (OAIS), welches Aufnahme, Sicherung und Verfügbarmachung der Daten definiert (vgl. Puhl et al. 2015, S. 30-32). Es besteht auch die Möglichkeit, diese Aufgaben an einen Dienstleister abzugeben. Einrichtungen sollten hierbei jedoch darauf achten, dass ein Service-Level-Agreement abgeschlossen wird, um alle erforderlichen Schritte sicherzustellen (vgl. Beucke 2010, S. 31-37).

3.2. Problematik Urheberrechte und Datenschutz.

Von der rechtlichen Seite her sind Digitalisate zum einen durch das Urheberrecht mit Fragen zum Copyright und Bildrechten betroffen, zum anderen durch Fragen zum Datenschutz- und Personenschutzrecht (vgl. Ludwig/Enke 2013, S. 55; vgl. Simukovic et al. 2014, S. 10f.).

Ein wichtiger Aspekt dabei ist, dass sich durch die elektronische Form von Digitalisaten eine andere Rechtslage ergibt als bei den analogen Originalen (vgl. Steinhauer 2017, S. 345): Lange Zeit waren Rechtsfragen bei der Nutzung von Quellen relativ unwichtig für die Geisteswissenschaften, da sich die Benutzung von Objekten mit Trägermedium rechtlich allein auf die Zugänglichkeit beschränkt. Da diese in Gedächtniseinrichtungen gesammelt und (zumindest der Wissenschaft) frei zugänglich gemacht werden, entsteht hier „ein

¹⁸ Auf die einzelnen Vor- und Nachteile der Verfahren wie z.B. Dateiverlust oder den notwendigen Wiederholungen kann an dieser Stelle leider nicht eingegangen werden.

¹⁹ Für einen idealtypischen Ablauf der Langzeitarchivierung und des Ingest siehe: Beucke 2010, S. 28f.

relativ unkomplizierter und von rechtlichen Vorgaben weitgehend freier Raum für wissenschaftliches Arbeiten“ (Steinhauer 2017, S. 346). Das Verbreitungsrecht des Nutzungsrechteinhabers am Objekt geht durch den Erschöpfungsgrundsatz beim Verkauf verloren, das Werk (nicht der Inhalt) steht der Einrichtung frei zur Verfügung, wobei Werk und Inhalt durch die Veröffentlichungsform eine Einheit bilden (vgl. Steinhauer 2017, S. 346).

Durch die Digitalisierung änderte sich dies jedoch, da die digitalen Objekte nun unabhängig von einem konkreten Trägermedium z.B. in Netzwerken gespeichert sind. Aus juristischer Sicht stellt die Digitalisierung daher mehr als einen Medienwandel dar (vgl. Steinhauer 2017, S. 345). „Mit dem Wegfall der Trägermedien rücken nun die Inhalte selbst in den Fokus“ (Steinhauer 2017, S. 346), was zu einer geänderten juristischen Situation führt. Nun wird direkt auf den digitalen Inhalt zugegriffen. Dieser kann, anders als analoge Texte, nicht nur auf vielen Geräten gleichzeitig gelesen werden, aus juristischer Sicht ist jedes Aufrufen des Objektes auf Grund der technischen Übertragung der Daten in den lokalen (temporären) Speicher des Nutzungsgerätes eine Vervielfältigung (vgl. Ludwig/Enke 2013, S. 75; vgl. Steinhauer 2017, S. 346) und damit ein Eingriff in die urheberrechtlichen Verwertungsrechte, welche im Kern aus den Rechten zur Vervielfältigung, Verbreitung, Ausstellung und öffentlichen Wiedergabe bestehen (vgl. §§15-18 UrhG). Ein Erschöpfungsgrundsatz besteht bei digitalen Objekten nicht, da diese immer wieder vervielfältigt werden können (vgl. Steinhauer 2017, S. 347). Für die Nutzung muss daher das erforderliche Nutzungsrecht vorhanden sein. Dieses wird durch den Urheber und die ihm zustehenden Verwertungsrechte vergeben und kann entweder als ausschließliches Nutzungsrecht auch den eigentlichen Urheber von der Nutzung ausschließen oder als einfaches Nutzungsrecht räumlich, zeitlich oder inhaltlich eingeschränkt sein (vgl. §31 UrhG).

Neben der Legitimierung durch Nutzungsrechte bilden Schrankenregelungen eine weitere Möglichkeit, urheberrechtlich geschützte Daten zu nutzen. So ist z.B. das Problem, dass technische Übertragungen bereits urheberrechtliche Vervielfältigungshandlungen sind, durch eine Schranke gelöst, da diese „vorübergehende Vervielfältigungshandlungen, die flüchtig oder begleitend sind und einen integralen und wesentlichen Teil eines technischen Verfahrens darstellen“ (§44a UrhG) erlaubt. Weitere Schranken für die Nutzung von Retrodigitalisaten betreffen die Bereiche Lehre und Forschung (vgl. §§60a-c UrhG) sowie, besonders wichtig für die Digital Humanities, Text und Data Mining (vgl. §60d UrhG). Bibliotheken im Speziellen und Gedächtniseinrichtungen im Allgemeinen werden in §§60e-f UrhG gesondert behandelt und erhalten das Recht „Werk aus ihrem Bestand oder ihrer Ausstellung für Zwecke der Zugänglichmachung, Indexierung, Katalogisierung, Erhaltung und Restaurierung vervielfältigen oder vervielfältigen lassen, auch mehrfach und mit technisch bedingten Änderungen“ (§60e Abs. 1 UrhG). Der Aspekt der technisch bedingten Änderungen ist zentrale Voraussetzung für die digitale Langzeitarchivierung, welche ansonsten eine unerlaubte Veränderung und Vervielfältigung (vgl. §§14, 16 und 39 UrhG) darstellen würden. Zudem erlaubt §60e Abs. 4 UrhG es, 10% von Werken sowie ganze Abbildungen, Zeitschriftenartikel, vergriffene und kurze Werke digital in den Räumen der Gedächtniseinrichtung zur Verfügung zu stellen. Nicht erlaubt ist damit jedoch die Verfügbarmachung von urheberrechtlich

geschützten Beständen der Gedächtniseinrichtung im Internet (vgl. Klimpel et al. 2017, S. 20). Diese kann nur durch Absprachen mit den Rechteinhabern erlaubt werden. Schranken bieten den Vorteil, dass sie ohne weitere Absprachen gelten. Dabei dürfen sie jedoch nicht die normale Auswertung eines Werkes beeinträchtigen (Drei-Stufen-Test). Eine Berufung auf Schranken für die Digitalisierung von analogen Werken, die durch den Rechteinhaber auch digital angeboten werden, ist daher nicht möglich (vgl. Steinhauer 2017, S. 349).

Diese Unterschiede in der Nutzung von analogen und digitalen Objekten fordert daher ein höheres Bewusstsein für die damit verbundenen Rechtsfragen. War bei physischen Objekten der Besitz zur Nutzung ausreichend, sind für Digitalisate Nutzungsrechte oder Schrankenregelungen notwendig (vgl. Steinhauer 2017, S. 347f.).

Das Urheberrecht an sich gilt für alle persönlich geschaffenen, wahrnehmbaren geistigen Leistungen, also Werke, die eine gewisse Schöpfungshöhe erreichen, bis 70 Jahre nach dem Tod des Urhebers, wobei der Urheber nur eine natürliche Person sein kann (vgl. §2 UrhG; vgl. Steinhauer 2017, S. 348). Der Urheber besitzt die unveräußerlichen Urheberpersönlichkeitsrechte (vgl. §§12-27 UrhG), welche die Grundlage der bereits beschriebenen Verwertungsrechte darstellen.

Für Forschungsdaten bedeutet dies, dass sie nicht immer die geistige Schöpfungshöhe erreichen und damit nicht immer urheberrechtlich geschützt sind. Maschinell erstellte und ausgegebene Messreihen, die ohne besondere manuelle Einstellungen eines Wissenschaftlers entstehen, sind daher nicht urheberrechtlich geschützt (vgl. Steinhauer 2017, S. 348). Dieser Sachverhalt lässt sich auch auf Digitalisate übertragen. Deren Ziel ist die möglichst genaue Wiedergabe der Vorlage ohne eine eigene künstlerische Leistung, weshalb das Urheberrecht hier nicht gilt (vgl. Klimpel et al. 2017, S. 38). Unter Umständen kann jedoch ein Leistungsschutzrecht für die Erstellung des Digitalisates entstehen. Voraussetzung dafür ist, dass die Reproduktion durch einen Fotografen oder Scan-Operator erfolgt, der die Einflussfaktoren und Einstellungen am Gerät manuell so einstellt, dass das beste Ergebnis erzielt wird. Es muss also eine Leistung erbracht werden, die geschützt werden soll. Einfache technische Reproduktionen wie Fotokopien, bei denen auch die Einstellungen des Scanners automatisch erfolgen, genießen dagegen keine Leistungsschutzrechte (vgl. Klimpel et al. 2017, S. 38; vgl. Steinhauer 2017, S. 347).

Leistungsschutzrechte schützen dabei als dem Urheberrecht verwandte Schutzrechte die Leistung, die erbracht wird, um ein Werk, nicht den Inhalt, zu erstellen. Sie sind 50 bis 70 Jahre nach der Erstveröffentlichung gültig und können auch juristischen Personen zustehen (vgl. Klimpel et al. 2017, S. 11f.). Bei Lichtbildern beträgt die Schutzfrist 50 Jahre (vgl. §72 UrhG; vgl. Steinhauer 2017, S. 347). Hinzu kommt als Sui-generis-Schutz der Schutz von Datenbanken (vgl. §§87a-e UrhG sowie vgl. Steinhauer 2017, S. 347f.), welcher auch für die Korpuserstellung von Bedeutung sein kann.

Insgesamt muss also nicht nur auf die Eigentumsrechte am Inhalt, also dem digitalisierten Original, sondern auch auf die Rechte am Digitalisat an sich sowie an möglichen digitalen Sammlungen geachtet werden (vgl. Czmiel et al. 2005, S. 157). Problematisch ist dies bei verwaisten Werken, also Werken, die auf Grund ihres Schöpfungszeitpunktes sicher oder

höchstwahrscheinlich noch unter dem Urheberrechtsschutz stehen, deren Rechteinhaber aber nicht bekannt ist (vgl. Steinhauer 2017, S. 349). Vor einer Digitalisierung muss hier sorgfältig nach möglichen Rechteinhabern gesucht und die Suche dokumentiert werden, sollte nachträglich eine Rechteinhaber festgestellt werden, so muss die Nutzung des jeweiligen Werkes beendet, bei Digitalisaten also die Zugänglichmachung verhindert werden (vgl. §§61-61c UrhG). Gerade im Hinblick auf eine gesicherte und persistente Verfügbarkeit von Forschungsdaten ist dies problematisch.

Auch wenn eine Einrichtung die vollen Verwertungsrechte zu den von ihr digitalisierten Beständen hat, so sind dennoch einige Einschränkungen zu beachten, wenn die Daten zugänglich gemacht werden sollen. So können zum einen persönlichkeits- oder datenschutzrechtliche Aspekte, auf welche später noch eingegangen wird, eine Rolle für die Veröffentlichung spielen (vgl. Diesterhöft 2014, S. 82). Zum anderen ist „darauf zu achten, dass keine Daten mit volksverhetzenden, pornografischen, ehrverletzenden oder gegen Bestimmungen des Patentrechts verstoßenden Inhalte gespeichert werden“ (Ludwig/Enke 2013, S. 81). Auf Aspekte bei der Nutzung dieser Daten, z.B. einem Nutzungszertifikat, welches bei Verstoß gegen die Nutzungsbedingungen entzogen werden kann (vgl. Rfll 2017, S. 19), kann an dieser Stelle nicht weiter eingegangen werden.

Werden Daten wie Digitalisate schließlich zugänglich gemacht, ist es wichtig, die Nutzungslizenz zu bestimmen. So werden klare Nutzungsrechte festgelegt, eine zentrale Voraussetzung, um sie für die Forschung zugänglich und für verschiedene Forschungsmethoden verfügbar zu machen (vgl. Schöch 2013, S. 2; vgl. Steinhauer 2017, S. 351).

Für die freie Nutzung der Daten im Sinne von Open Access, welche für die wissenschaftliche Nutzung wichtig ist und daher von Gedächtniseinrichtungen unterstützt werden sollte, stehen dabei verschiedene Lizenzarten zur Verfügung. Die in Gedächtniseinrichtungen am weitesten verbreitete Gruppe sind die Creative Commons Lizenzen, welche eine Art Baukastensystem für verschiedene Nutzungsabstufungen geben, bestehend aus einer Namensnennung (CC-BY), einer Weitergabe unter gleichen Bedingungen (CC-SA), dem Verbot von Veränderungen (CC-ND) oder dem Verbot der kommerziellen Nutzung (CC-NC). Da eine freie Lizenz nach der Veröffentlichung der Daten nicht mehr geändert werden kann, haben diese funktional den Charakter des Erschöpfungsgrundsatzes bei körperlichen Werken (vgl. Steinhauer 2017, S. 351).

Da es den Managementaufwand innerhalb einer Infrastruktur erhöht, wenn Inhalte mit unterschiedlichen Lizenzen oder gar Lizenzarten verknüpft sind, sollte bei der Vergabe hier auf Einheitlichkeit geachtet werden (vgl. Puhl et al. 2015, S. 38). Dies erleichtert auch die Pflege, wenn Lizenzen im Nachhinein, z.B. nach dem Ablauf von Sperrfristen, angepasst werden müssen (vgl. Cremer et al. 2018, S. 155).

Für die Nutzung in den Digital Humanities sind Lizenzen besonders wichtig, da es hier eben nicht nur um den Zugang zu Digitalisaten oder Text geht, sondern auch um deren

Weiterverarbeitung. Rechtsfragen stellen hier also ein wichtiges Feld dar (vgl. Steinhauer 2017, S. 351).

Verfahren wie Text oder Data Mining stellen als Arbeitsprozesse zwar Vervielfältigungen dar, sind inzwischen jedoch durch eine Schranke im Urheberrecht für die wissenschaftliche Nutzung erlaubt. Daraus resultierende Korpora müssen zwar vom Nutzer gelöscht, dürfen zur Archivierung aber in Gedächtniseinrichtungen gegeben werden (vgl. §60d UrhG).

Damit stellt die Analyse von Daten zwar kein urheberrechtliches Problem dar, Voraussetzung ist aber, dass die untersuchten Objekte digital vorliegen und verfügbar sind. Und da die Verfügbarmachung von geschützten Objekten wie bereits beschreiben in Gedächtniseinrichtungen gerade bei großen Werken nur zum Teil möglich ist (vgl. §§60c-f UrhG), bleibt hierbei oftmals nur der Erwerb der Nutzungsrechte oder eine Analyse von unvollständigen Korpora übrig, was die wissenschaftliche Nutzbarkeit stark einschränkt (vgl. Kaden/Kleineberg 2016, S. 51; vgl. Steinhauer 2017, S. 348).

Neben urheberrechtlichen Problemen sind Persönlichkeitsrechte und Datenschutz weitere Aspekte, auf die bei der Verfügbarmachung von Digitalisaten geachtet werden muss.

Das Persönlichkeitsrecht lässt sich aus dem Grundgesetz Art. 1 Abs. 1 zur Menschenwürde in Verbindung mit Art. 2 Abs. 1 zur allgemeinen Handlungsfreiheit ableiten. Es beinhaltet die informationelle Selbstbestimmung, also wie man sich selber gegenüber Dritten darstellen möchte, da dies für die Entwicklung der eigenen Persönlichkeit wichtig ist. Für jede Art von Erhebung, Verarbeitung und Weitergabe von personenbezogenen Daten muss also die jeweils betroffene Person befragt werden (vgl. Diesterhöft 2014, S. 55f.), wobei personenbezogene Daten alle Daten sind, deren Informationen sich klar der jeweiligen Person zuordnen lassen (vgl. Klimpel et al. 2017, S. 34).

Allerdings kann auch dieses Recht durch Schranken begrenzt werden. Im Falle des Grundrechts auf informationelle Selbstbestimmung liegt die Grundlage hierfür im Grundrecht der Informations- und Wissenschaftsfreiheit (vgl. Art. 5 Abs. 3 GG), wobei stets eine Abwägung zwischen dem Persönlichkeitsrecht und der Wissenschaftsfreiheit nötig ist (vgl. Diesterhöft 2014, S. 65). Personenbezogene Daten dürfen daher für wissenschaftliche Zwecke verarbeitet werden, müssen aber wenn möglich anonymisiert und am Ende der Forschung gelöscht werden (vgl. Ludwig/Enke 2013, S. 56). Öffentlich arbeitende Archive dürfen entsprechende Daten, auch ohne Einwilligung der betroffenen Person (vgl. §27 BDSG), speichern und werden auch in weiteren Punkten wie den Anforderungen an die Zweckbindung, die Verarbeitung oder die Begrenzung der Speicherdauer bevorzugt (vgl. Klimpel et al. 2017, S. 34f.).

Damit ist zwar die wissenschaftliche Nutzung der entsprechenden Daten in den jeweiligen Einrichtungen, besonders Archiven, ermöglicht, nicht jedoch eine allgemeine Veröffentlichung, wie sie im Internet stattfinden würde. Hier ist es weiterhin nötig, die Erlaubnis der betroffenen Person einzuholen. Dies gilt auch bei digitalisierten Werken, die bereits veröffentlicht sind, da die Verfügbarmachung im Internet eine erneute Veröffentlichung und damit eine neue Beeinträchtigung des Persönlichkeitsrechts darstellt, besonders, da mögliche Einwilligungen unter den neuen

Rezeptionsbedingungen des Internets möglicherweise nicht mehr gelten (vgl. Diesterhöft 2014, S. 81). Neue Rezeptionsbedingungen des Internets sind dabei vor allem der Wegfall einer überprüfbaren Zweckbindung, welche bei freiem Internetzugang nicht möglich ist, sowie die Volltextsuche. Durch diese kann nicht nur gezielt nach Personen gesucht werden, sondern es sind auch Personen, die nur am Rande erwähnt und sonst nicht elektronisch erfasst werden, leicht auffindbar (vgl. Diesterhöft 2014, S. 74 sowie S. 76).

Werden diese Digitalisate nun durch Gedächtniseinrichtungen erstellt und zugänglich gemacht, so sind diese, und nicht wie bei analogen Werken der Verlag oder Autor, für die Veröffentlichung verantwortlich und damit auch für die Grundrechtseingriffe durch die neuen Rezeptionsbedingungen haftbar (vgl. Diesterhöft 2014, S. 67f. sowie S. 79f.).

Bei der Digitalisierung von personenbezogenen Daten sollte daher eine Prüfung der Informationen erfolgen. Bei Unklarheit sollte dann das Verblassen von staatlichen Schutzfristen bzw. des (postmortalen) Persönlichkeitsschutzes abgewartet werden (vgl. Diesterhöft 2014, S. 77), welche im Bundesarchivgesetz auf 30 Jahre nach Entstehen der Unterlagen (allgemeine Schutzfrist), zehn Jahre nach dem Tod der betroffenen Person oder, wenn nicht feststellbar, 100 Jahre nach Geburt der Person, festgelegt ist (vgl. §11 BArchG).

Urheberrechtlich unproblematisch sind historische Bestände bis zum Beginn des 20. Jahrhunderts (vgl. Diesterhöft 2014, S. 54), jedoch ist das Persönlichkeitsrecht in Teilen auch auf Verstorbene anzuwenden und daher hier zu beachten.

Das allgemeine Persönlichkeitsrecht erlischt mit dem Tod, da seine Funktion des Schutzes der Fortentwicklung der eigenen Persönlichkeit nach dem Tod nicht mehr besteht. Bestehen bleibt dagegen die unantastbare Menschenwürde (Art. 1 Abs. 1 GG) des Verstorbenen und damit ein Schutz des sittlichen, persönlichen und sozialen Geltungswertes sowie der Achtungsanspruch des Verstorbenen, welcher vor Herabsetzung, Entstellung, Verunglimpfung und Beleidigung schützt (vgl. Diesterhöft 2014, S. 60; vgl. Klimpel et al. 2017, S. 33).

Bei der Digitalisierung muss daher zwar jeweils entschieden werden, ob der Achtungsanspruch eines Verstorbenen durch die Verfügbarmachung persönlicher Daten in Digitalisaten beeinträchtigt ist (vgl. Diesterhöft 2014, S. 62), doch die reine Wiedergabe von Daten, wie sie in den meisten Dokumenten der Fall und von den neutralen Gedächtniseinrichtungen angestrebt ist, sollte unproblematisch sein (vgl. Klimpel et al. 2017, S. 33).

Schwieriger ist dies, wenn Daten von anderen, möglicherweise noch lebenden Personen, betroffen sind. Das betrifft zum einen Briefwechsel, bei denen zudem die Urheberrechte aller Beteiligten zu beachten sind, zum anderen aber auch Tagebücher oder Berichte Verstorbener, die Informationen über noch Lebende enthalten. Hier sollte überprüft werden, ob der Verstorbene diese Daten hätte selber veröffentlichen dürfen, was auch Bereiche des Presse- und Ehrschutzrechtes beinhaltet (vgl. Klimpel et al. 2017, S. 33).

Zu berücksichtigen sind zudem die Interessen der Angehörigen. Nach dem Tod nimmt das Schutzbedürfnis daher in dem Maße ab, wie die Erinnerung an den Verstorbenen sowie das Interesse an der Nichtverfälschung seines Lebensbildes abnimmt (vgl.

Diesterhöft 2014, S. 61). Dies betrifft nicht nur das Ansehen des Verstorbenen (und damit mittelbar der Angehörigen selber) (vgl. Diesterhöft 2014, S. 61), sondern auch vermögenswerte Interessen wie z.B. Rechte an Darstellungen des Toten (auch noch aus Lebzeiten), welche nach Kunsturhebergesetz noch zehn Jahre nach dem Tod gültig sind (vgl. §22 KunstUrhG). Zwar gilt diese Frist nur für Bilder, nicht für andere Darstellungsformen wie Tonaufzeichnungen, doch allgemein zeigt sich, dass Ansprüche hierbei vor allem für die kommerzielle Nutzung relevant sind, diese bei Gedächtniseinrichtungen aber nicht gegeben ist (vgl. Klimpel et al. 2017, S. 34).

Die Schutzfristen nehmen also kontinuierlich ab, mit genügend Zeit dürfen selbst sensibelste Informationen veröffentlicht werden (vgl. Diesterhöft 2014, S. 61), solange der allgemeine Achtungsanspruch bestehen bleibt und die Informationen nicht Rückschlüsse auf noch lebende Personen erlauben oder diese gar beeinträchtigen können.

4. Digitalisate in Kulturerbeeinrichtungen.

Die Kernaufgaben von Gedächtniseinrichtungen bestehen aus den Bereichen Sammeln, Beschreiben, Aufbewahren und zugänglich machen. Dies betrifft nicht nur analoge Werke, sondern auch, und zwar unabhängig vom Fachgebiet, Forschungsdaten, welche idealerweise gesammelt, aufbereitet, annotiert und erschlossen werden, um sie nachweisen zu können (vgl. Kaden/Kleineberg 2016, S. 49). Für die Nutzer sind sie daher Ansprechpartner für die Verwaltung und Archivierung (vgl. Kaden/Kleineberg 2016, S. 48), deren inhaltliche Bearbeitung bei der Erfassung endet, während die inhaltliche Auswertung Aufgabe der Nutzer ist (vgl. Münster et al. 2018, S. 379; vgl. Voß 2013, S. 8). Konkret lassen sich diese Aufgaben auch auf die Digitalisate als Forschungsdaten übertragen (vgl. Martin 2013, S. 20). Der Sonderfall bei Digitalisaten besteht jedoch darin, dass sie anders als andere Forschungsdaten in der Regel in den Gedächtnisinstitutionen erzeugt werden. Dies führt zu einigen Unterschieden im Vergleich zu anderen Forschungsdaten: Besteht bei sonstigen Forschungsdaten oft das Problem, dass, wenn nicht von der Entstehungsinstitution bereitgestellt, nur unklare Infrastrukturen zur Sicherung der Daten vorliegen (vgl. Sahle/Kronenwett 2013, S. 79f.), so ist dies bei Retrodigitalisaten kein Problem, da hier üblicherweise die Einrichtung, die die Digitalisierung durchgeführt hat, auch die Digitalisate speichert, um Originale zu schonen und Investitionen zu bewahren. In der Regel findet die Digitalisierung dabei in den Gedächtniseinrichtungen statt, welche auch die Originale besitzen. Eine Überprüfung der Integrität ist daher z.B. nur teilweise nötig, da keine Fremddaten verarbeitet, sondern nur Daten der eigenen Einrichtung angeboten werden. Bei diesen kann die Qualität und Richtigkeit direkt kontrolliert werden. Die Verfügbarmachung durch Gedächtniseinrichtungen betrifft auch die Metadatenaufbereitung, für die bei anderen Forschungsdaten oftmals die finanziellen oder zeitlichen Ressourcen fehlen (vgl. Beucke 2010, S. 6), sowie die Nutzungsrechte. Schränken Forscher diese bei eigenen Daten teilweise ein (vgl. Kaden 2018, S. 2-5; vgl. Kindling et al. 2013, S. 56), so ist die Zugänglichmachung der Daten durch Gedächtniseinrichtungen eine zentrale Aufgabe, die Digitalisate sind genuin für die Veröffentlichung bestimmt.

Und auch Probleme der sehr heterogenen Standards und Dateiformate (vgl. Beucke 2010, S. 6; vgl. Simukovic et al. 2014, S. 14) sind nicht gegeben, da die jeweilige Einrichtung sich zumindest an eigene einheitliche Standards halten sollte oder idealerweise die DFG-Praxisregeln zur Digitalisierung befolgt. Das hat auch zur Folge, dass sich die Kosten für Langzeitarchivierung und Speicher besser abschätzen lassen, was sonst auf Grund der unterschiedlichen Forschungsdaten nur schwer möglich ist (vgl. Beucke 2010, S. 26; vgl. Pempe 2012, S. 148). Allgemein ist die nachhaltige Pflege der Daten in den langfristigen Gedächtniseinrichtungen, anders als bei projektbezogenen Forschungsdatensammlungen, sichergestellt (vgl. Rapp 2018, S. 255).

Das die Gedächtniseinrichtungen ihre Bestände digitalisieren und verfügbar machen, wird dabei von der Wissenschaft und der Politik bzw. Geldgebern gefordert, um die Bestände einer möglichst breiten Öffentlichkeit zugänglich zu machen (vgl. KII 2011, S. 143).

Trotz verschiedener Digitalisierungsprojekte ist bisher nur ein geringer Teil des kulturellen Erbes in Deutschland digitalisiert (vgl. Fickers 2014, S. 27; vgl. Klaffki et al. 2018, S. 6-12), wobei Schwerpunkte auf schriftlichem Kulturgut, Zeitungen oder Fotografien liegen, während Objekte oder umfangreiche Akten kaum digitalisiert sind (vgl. Fickers 2014, S. 27; vgl. KII 2011, S. 28; vgl. Nauta/van den Heuvel 2015, S. 4).

Die Digitalisierung, auch als Teil der Bestandserhaltung, erfolgt dabei durch die besitzenden Gedächtniseinrichtungen. Zwar werden oftmals Dienstleister mit der Digitalisierung beauftragt, auf Dauer ist aber, gerade für große Einrichtungen, der Aufbau eines eigenen Digitalisierungszentrums günstiger (vgl. Czmiel et al. 2005, S. 23f.). Die Verknüpfung der digitalen Objekte mit Metadaten, Sammlungen und Projekten kann so direkt stattfinden (vgl. Czmiel et al. 2005, S. 153), zudem können kleine Einrichtungen durch die Übernahme von Digitalisierungsprojekten unterstützt und die Zusammenarbeit unter öffentlich geförderten Einrichtungen gestärkt werden (vgl. KII 2011, S. 38). Für aufwändige Aufgaben wie Transkriptionen kann zudem auf Crowdsourcing statt auf Dienstleister zurückgegriffen werden, womit zugleich die Sichtbarkeit der Institution erhöht wird (vgl. Schubert 2018, S. 239).

Das Pflegen von digitalen Sammlungen ist eine wichtige Aufgabe von Gedächtniseinrichtungen, dient es nicht nur dem eigenen Bestandsmanagement (vgl. Fuhr et al. 2001, S. 188), sondern Gedächtniseinrichtungen stellen so auch im Internet eine verlässliche und qualitätsgesicherte Informationsquelle für Forschung und interessierte Öffentlichkeit und damit einen Gegenpol gegen „zahllose Laienangebote, häufig sogar unseriöse und fragwürdige Angebote im Netz“ (Rapp 2018, S. 251) dar. Die Nutzer vertrauen diesen etablierten Einrichtungen bei der Sicherung, Aufbereitung und Zugänglichmachung der Daten ebenso wie bei ihrer inhaltlichen Richtigkeit (vgl. Baru 2007, S. 116).²⁰

Bedeutung kommt den digitalen Angeboten zudem zu, da inzwischen die meisten Objekte einer Einrichtung über ihre digitale Sammlung und nicht mehr vor Ort genutzt werden (vgl. Nauta/van den Heuvel 2015, S. 5). Damit ist zum einen auf eine ansprechende Präsentation und Funktionalität zu achten, es können zum anderen aber auch Objekte, die nicht in den regulären Ausstellungen Platz finden, angemessen publik gemacht werden. Durch entsprechende Auswertungen erhält die Einrichtung zudem Informationen über die Interessen der Nutzer, welche sich auch auf andere Geschäftsbereiche (Ausstellungsplanung, Neuankäufe, ...) auswirken können (vgl. Ullmann 2016, S. 34 sowie 37f.).

Zur digitalen Sammlung gehört dabei neben den Informationen an sich auch das System mit Benutzeroberfläche und Funktionen sowie die Infrastruktur und das Personal dahinter (vgl. Czmiel et al. 2005, S. 153; vgl. Fuhr et al. 2001, S. 191). Die Vorteile von Online-Sammlungen für die Nutzer wurden bereits beschrieben²¹, hinzu kommt, dass die

²⁰ Siehe hierzu auch Abschnitt 5.4.2. Auswertung der Interviews - Nutzer.

²¹ Siehe Abschnitt 3. Digitalisate als Forschungsdaten.

Institutionen hier als Data Provider fungieren und Aggregatoren oder Plattformen beliefern können, welche gesammelt Zugang zu Daten aus verschiedenen Einrichtungen bieten (vgl. Münster et al. 2018, S. 377-379). Portale bieten dabei für die Institution den Vorteil, dass Informationen vernetzt und somit die Sichtbarkeit erhöht wird. Zudem sind Portale in der Regel besser durch Suchmaschinen indexiert als einzelne Angebote (vgl. Schubert 2018, S. 225). Einrichtungen können so aktiv an der Verbreitung von strukturierten und qualitätsgesicherten Forschungsdaten teilhaben und zu einer Vernetzung unter den Institutionen sowie von Institutionen und Wissenschaft beitragen (vgl. Allianzinitiative 2018, S. 6). Hinzu kommt, dass Forscher eher dazu bereit sind, gefundene Informationen zu teilen, wenn sie leicht Zugang zu diesen bekommen haben (vgl. Münster et al. 2018, S. 372).

Um dies zu erreichen, sollten die Objekte in digitalen Sammlungen genauso wie andere Forschungsdaten auch ausreichend mit Angaben zu Umfang, Format, Herkunft und Lizenz ausgezeichnet (vgl. Czmil et al. 2005, S. 154) und in einer hohen, möglichst breit nutzbaren Auflösung angeboten werden (vgl. Münster et al. 2018, S. 379). Begriffe wie Open Access oder Open Data tauchen in diesem Zusammenhang selten auf, viele öffentlich geförderte Einrichtungen und Projekte²² setzen dies aber praktisch um, indem die Digitalisate und Metadaten technisch und finanziell offen über das Internet verfügbar sind. Durch Transkriptionen alter Handschriften wird eine weitere Hürde der Nutzung abgebaut, um die Dokumente auch inhaltlich zu öffnen.

Da durch eine Öffnung der Angebote nicht mehr nur Expertennutzer bzw. Wissenschaftler angesprochen werden sollen, sondern auch die breite Öffentlichkeit, ist es wichtig darauf zu achten, auch deren Anforderungen zu berücksichtigen. Neben Filtern und Expertensuchen sollten daher auch einfaches Suchen und Browsen Möglichkeiten des Zugangs darstellen, da sonst auch gut aufbereitete Daten kaum genutzt werden (vgl. Given/Willson 2018, S. 809; vgl. Münster et al. 2018, S. 380). Auch eine gute Erreichbarkeit über die gängigen Web-Suchmaschinen, gerade unter Berücksichtigung der zunehmenden Personalisierung dieser, ist ein wichtiger Aspekt, um die digitalen Sammlungen und damit die Objekte auffindbar zu machen (vgl. Haber 2012, S. 8).

²² Beispiele hierfür sind das Projekt 100(0) Schlüsseldokumente zur deutschen Geschichte der Bayerischen Staatsbibliothek (1000 Schlüsseldokumente zur deutschen Geschichte im 20. Jahrhundert, https://www.1000dokumente.de/index.html?c=1000_dokumente_de&viewmode=0&l=de (abgerufen am 20.07.2019, 14:50)), die Digitalisierte Sammlung der Staatsbibliothek zu Berlin (Staatsbibliothek zu Berlin. Digitalisierte Sammlungen, <https://digital.staatsbibliothek-berlin.de/> (abgerufen am 20.06.2019, 23:34)), verschiedene digitalisierte Archivalien des Geheimen Staatsarchivs Preußischer Kulturbesitz (Geheimes Staatsarchiv. Digitalisierte Archivalien, https://www.gsta.spk-berlin.de/digitalisierte_archivalien_1612.html (abgerufen am 20.07.2019, 14:47)) oder die Sammlung Online des Stadtmuseums Berlin (Stadtmuseum Berlin. Sammlung Online, <https://sammlung-online.stadtmuseum.de/> (abgerufen am 24.05.2019, 13:45)).

5. Untersuchung von Digitalisaten in Kulturerbeeinrichtungen.

5.1. Vorstellung der Interviewpartner und Institutionen.

Um einen Abgleich zwischen Anforderungen der Nutzer und Möglichkeiten der Institutionen zu ermöglichen, werden Interviews mit beiden Gruppen geführt. Da diese relativ ausführlich auf verschiedene Aspekte der Auswahl, Herstellung, Verwaltung und Nutzung von Digitalisaten und den zugehörigen Darstellungsplattformen eingehen sollen, fanden jeweils nur drei Interviews pro Gruppe statt. Die Interviews fanden in der Regel mit einem, in zwei Ausnahmen auch mit zwei Interviewpartnern statt.

Auf Seiten der Gedächtniseinrichtungen wurden jeweils die für die (Retro-)Digitalisierung entscheidenden Mitarbeiter in der Stiftung Stadtmuseum Berlin, der Staatsbibliothek zu Berlin Preußischer Kulturbesitz und der Universitätsbibliothek der Humboldt-Universität zu Berlin befragt.

Die Einrichtungen sind zwar alle groß bis sehr groß, haben jedoch unterschiedliche Ausrichtungen von internationalen, über nationale bis hin zu regionalen Schwerpunkten. Wichtig ist zudem, dass alle der hier untersuchten Institutionen noch aktiv ihre Objekte digitalisieren, der Prozess der Retrodigitalisierung also noch nicht abgeschlossen ist, und dies auch im Haus geschieht. Zudem werden die Digitalisate auf eigenen Websites und nicht nur auf extern gehosteten Portalen angeboten. Bei kleineren Institutionen sind weder die interne Digitalisierung noch die eigenständige Präsentation unbedingt gegeben, so dass diese nicht geeignet für Interviews wären.

Die Stiftung Stadtmuseum Berlin wurde 1995 als Fusion des Märkischen Museums (1874 als Märkisches Provinzial-Museum gegründet), des Berlin Museums und 14 weiterer kleinerer Museen aus der gesamten Stadt gegründet und besteht heute aus fünf Standorten.²³ Thematisch sammelt das Museum dabei im Bereich der Berliner Geschichte und Kultur und besitzt ca. 4,5 Mio. Objekte, unterteilt in verschiedene Sammlungen zur Stadtgeschichte, Alltagskultur, Kunstgewerbe oder Theater, Musik und Literatur.²⁴ Die Digitalisierung findet vor allem projektweise, aber auch für Ausstellungen oder auf Nutzeranfrage statt, digitale Objekte werden in der Sammlung Online sichtbar gemacht.²⁵ Interviewpartner aus dem Stadtmuseum Berlin waren Frau Sabine Weller (Leiterin Fachteam Zentrale Dokumentation) und Herr Sebastian Ruff (Leiter eCulture).

²³ Vgl. Stadtmuseum Berlin. Geschichte der Stiftung Stadtmuseum Berlin, <https://www.stadtmuseum.de/geschichte-der-stiftung-stadtmuseum-berlin> (abgerufen am 25.07.2019, 23:40) sowie Stadtmuseum Berlin. Unsere Museen, <https://www.stadtmuseum.de/unsere-museen> (abgerufen am 25.07.2019, 23:40).

²⁴ Vgl. Stadtmuseum Berlin. Sammlungen, <https://www.stadtmuseum.de/sammlungen> (abgerufen am 25.07.2019, 23:40).

²⁵ Stadtmuseum Berlin. Sammlung Online, <https://sammlung-online.stadtmuseum.de/> (abgerufen am 24.05.2019, 13:45).

Die Staatsbibliothek zu Berlin, deren Ursprünge bis auf die 1661 gegründete kurfürstliche Bibliothek zurückgehen, ist eine der größten wissenschaftlichen Universalbibliotheken Deutschlands und Teil der Stiftung Preußischer Kulturbesitz. Ihrer Größe und der Bedeutung des Bestandes entsprechend ist sie Teil verschiedener nationaler wie internationaler Netzwerke, bspw. durch das Betreuen verschiedener FIDs, der Sammlung Deutscher Drucke 1871-1912 oder der Teilhabe an Europeana und der World Digital Library. Zum Bestand der Staatsbibliothek gehören über 11,5 Mio. Bücher (darunter ca. 200.000 Rara und 4600 Inkunabeln) sowie ca. 2,2 Mio. weitere Druckwerke oder Handschriften. Einen großen Teil der Sammlung stellen die historischen Bestände dar, wobei viele der Objekte unikal sind.²⁶

Zur Verfügbarmachung und Sicherung des großen historischen Bestandes sowie Koordination verschiedenster Digitalisierungsprojekte wurde im Jahr 2010 ein Digitalisierungszentrum eingerichtet. Die Aufgaben bestehen vor allem in der Koordination der Digitalisierung zwischen den beteiligten Stellen (den Bestandsabteilungen, der Restaurierung, dem Digitalisierungszentrum an sich sowie der Abteilung Informations- und Datenmanagement) sowie der eigentlichen Digitalisierung von Objekten im Rahmen von Projekten, der Schutz- und Sicherheitsdigitalisierung von besonders bedeutenden oder gefährdeten Beständen sowie der Digitalisierung im Auftrag von Nutzern (Digitizing on demand (DoD)).²⁷ Dazu stehen im Digitalisierungszentrum mehr als 15 verschiedene Scanner für unterschiedliche Formate (Bücher, Karten, Microfilme) sowie eine eigene Fotoabteilung zur Verfügung. Die digitalisierten Werke werden, soweit urheberrechtlich möglich, anschließend in den Digitalen Sammlungen der Staatsbibliothek im Internet frei verfügbar gemacht.²⁸

Als Interviewpartner standen von Seiten der Staatsbibliothek Frau Renate Müller (Leiterin Referat 2 Buchbearbeitung, Abteilung II A Bestandsaufbau) und Herr Thorsten Siegmann (Leiter Referat 3 Digitalisierung und Druckerei sowie Leitstelle Digitale Bibliothek, Abteilung II F Bestandserhaltung und Digitalisierung) zur Verfügung.

Die Universitätsbibliothek entstand 1831 im Zuge der Gründung der Berliner Universität und deckt heute thematisch alle Fachbereiche der Humboldt-Universität ab. Der Bestand umfasst ca. sechs Mio. gedruckte Medien sowie Zugang zu mehr als 200.000 digitalen Objekten. Die Universitätsbibliothek hat heute zwölf Standorte und ist eine der am meisten besuchten wissenschaftlichen Bibliotheken Deutschlands.²⁹ Die Altbestände der Bibliothek bestehen vor allem aus historischen Drucken und Gelehrtenbibliotheken oder -nachsätzen, aber auch Vorlesungsmitschriften oder thematischen Spezialsammlungen.³⁰

²⁶ Vgl. Staatsbibliothek zu Berlin. Porträt der Staatsbibliothek, <https://staatsbibliothek-berlin.de/die-staatsbibliothek/portraet/> (abgerufen am 25.07.2019, 20:30); vgl. Staatsbibliothek zu Berlin. Zahlen und Fakten, <https://staatsbibliothek-berlin.de/die-staatsbibliothek/zahlen-und-fakten/> (abgerufen am 25.07.2019, 20:30).

²⁷ Vgl. Staatsbibliothek zu Berlin. Bestandserhaltung und Digitalisierung, <https://staatsbibliothek-berlin.de/die-staatsbibliothek/abteilungen/bestandserhaltung/digitalisierung/> (abgerufen am 25.07.2019, 20:40).

²⁸ Staatsbibliothek zu Berlin. Digitalisierte Sammlungen, (abgerufen am 20.06.2019, 23:34).

²⁹ Vgl. Universitätsbibliothek der HU Berlin. Geschichte der Universitätsbibliothek, <https://www.ub.hu-berlin.de/de/ueber-uns/geschichte> (abgerufen am 25.07.2019, 23:23); vgl. Universitätsbibliothek der HU Berlin. Profil, <https://www.ub.hu-berlin.de/de/ueber-uns/profil> (abgerufen am 25.07.2019, 23:23).

³⁰ Vgl. Universitätsbibliothek der HU Berlin. Digitale Sammlungen browsen, <https://www.digi-hub.de/viewer/browse/> (abgerufen am 25.07.2019, 23:30); vgl. Universitätsbibliothek der HU Berlin.

Da diese digitalisiert und frei im Internet zugänglich gemacht werden, gehört die Digitale Sammlung der Universitätsbibliothek zum Bereich der Historischen Sammlungen.³¹

Interviewt wurde Frau Dr. Yong-Mi Rauch, die Leiterin der Abteilung Historische Sammlungen.

Die für die Interviews befragten Nutzer stammen alle aus der akademischen Geschichtsforschung. Dies ist insofern wichtig, da der Forschungsdatenbegriff bereits innerhalb der einzelnen Fachdisziplinen unklar ist (vgl. Cremer et al. 2018, S. 142), eine Ausweitung auf andere Bereiche wie die Populärwissenschaft oder Laienforschung diesen daher noch umfassender gestalten würde. Statt dessen soll sich hier auf einen akademischen Forschungsdatenbegriff konzentriert, dabei aber innerhalb der Disziplin ein möglichst breites Spektrum untersucht werden. In der Geschichtswissenschaft bietet sich dabei eine Verteilung über die verschiedenen Epochen an.

Der erste interviewte Nutzer möchte aus persönlichen Gründen anonym bleiben und wird daher im Weiteren als Nutzer 1 bezeichnet. Dieser ist Professor der Geschichtswissenschaft mit einem besonderen Schwerpunkt auf der (Frühen) Neuzeit. Aus gesundheitlichen Gründen kann Nutzer 1 nur eine begrenzte Zeit am Stück vor dem Bildschirm verbringen. Für die Auswertung ist dies ein weiterer Aspekt bei der Arbeit mit Digitalisaten, der berücksichtigt werden sollte.

Auch der zweite interviewte Nutzer möchte lieber anonym bleiben und wird im Folgenden daher mit Nutzer 2 benannt. Dieser ist Kunsthistoriker mit dem Schwerpunkt Denkmalpflege. Der relevante Zeitraum liegt im 18. und 19. Jahrhundert, wobei vereinzelt auch ältere oder neuere Akten, Pläne oder Abbildungen von Interesse sein können.

Als dritter Nutzer wurde Herr Dr. Andreas Kohring interviewt. Dieser ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Alte Geschichte an der Humboldt-Universität.³² Der Epochenschwerpunkt der Antike ist dabei besonders interessant, da hier wiederum andere Anforderungen an Digitalisate und Forschungsdaten gelten als in den späteren Epochen mit mehr überlieferten Objekten.

Herzlichen Dank an alle Interviewten für ihre Unterstützung.

5.2. Methodik der Interviews.

Ziel der Interviews war es herauszufinden, warum verschiedene Aspekte der Digitalisierung in Gedächtniseinrichtungen und zugehörige Entscheidungen jeweils in ihrer aktuellen Form ausgestaltet wurden bzw. wie Nutzung und Anforderungen an Digitalisaten

Historische Spezielsammlungen der Bibliothek, <https://www.ub.hu-berlin.de/de/literatur-suchen/sammlungen/sammlungen-sonderbestaende> (abgerufen am 25.07.2019, 23:30).

³¹ Universitätsbibliothek der HU Berlin. Digitale Sammlungen, <https://www.digi-hub.de/viewer/> (abgerufen am 14.06.2019, 12:30).

³² Vgl. HU Berlin. Institut für Geschichtswissenschaften. Dr. Andreas Kohring, <https://www.geschichte.hu-berlin.de/de/bereiche-und-lehrstuehle/alte-geschichte/alte-geschichte/personen/kohring> (abgerufen am 26.07.2019, 00:25).

aus Sicht der Nutzer sich zusammensetzen. Da es hierbei um Entscheidungen und Sichtweisen der jeweils Befragten geht, eignete sich eine qualitative Einzelbefragung besonders für die Untersuchung (vgl. Werner 2013, S. 129), zumal die Interviewpartner im Sinne eines Experteninterviews nicht nur Expertise zum Thema haben, sondern auch die Möglichkeit, ihre Entscheidungen in die Praxis umzusetzen (vgl. Werner 2013, S. 143).

Als Interviewmethode wurde hier ein teilstrukturiertes Experteninterview eingesetzt, eine Unterform der Leitfadeninterviews (vgl. Werner 2013, S. 142). Kernidee des teilstrukturierten Leitfadeninterviews ist die Orientierung an einem vorher erstellten Fragenkatalog, wobei die Fragen dennoch an die Situation angepasst oder durch nicht geplante Nachfragen ergänzt werden können (vgl. Werner 2013, S. 130-133). Die Reihenfolge der Fragen bewegt sich eigentlich vom Allgemeinen zum Speziellen (vgl. Werner 2013, S. 132), die Interviewfragen hier richteten sich jedoch nach den Aufgaben des Forschungsdatenkreislaufs (Institutionen) bzw. eines schematischen Arbeitsablaufs (Nutzer), damit nicht zwischen verschiedenen Arbeitsbereichen gesprungen werden musste. Innerhalb jedes Aspektes wurden die Fragen dann aber möglichst immer spezieller. Einige Fragen wurden sowohl Nutzern als auch Institutionen gestellt, um deren jeweiligen Standpunkte direkt vergleichen zu können.³³

Da es bei den Interviews nur um die inhaltlichen Aussagen geht, sind auch nur diese in den Transkriptionen der Interviews enthalten. Hinweise auf Unterbrechungen oder Reaktionen der Interviewpartner sind nicht nötig, da diese nicht für den Untersuchungsgegenstand (Nutzungsweisen, Anforderungen etc.) relevant sind (vgl. Werner 2013, S. 135).

Bei der Auswertung wurden, ähnlich dem Typisierungsmodell (vgl. Werner 2013, S. 138f.), die Aussagen der Interviewpartner zu den einzelnen Fragen verglichen und damit deren (idealen und realen) Standpunkte für die Teilbereiche des Forschungsdatenkreislaufs bzw. Arbeitsablaufs erarbeitet. Ausgehend von den Positionen der Nutzer fand anschließend ein Abgleich statt mit den Möglichkeiten und Positionen der Einrichtungen, wobei auch die Evaluation derer Websites Berücksichtigung fand.

Nach Fertigstellung der Interviewleitfäden fanden zwei Pre-Tests statt, um die Verständlichkeit der Fragen und Dauer des Interviews abschätzen zu können. Einige Fragen konnten dadurch noch umformuliert werden, um sie verständlicher zu machen.³⁴

Für die Umsetzung der Interviews wurde den Interviewpartnern ca. eine Woche vor dem Interview der Fragebogen zur Verfügung gestellt. Während der Interviews wurde, mit Einverständnis der Befragten, das Gespräch aufgezeichnet, so dass die Konzentration auf das Gespräch und nicht das Anfertigen von Notizen etc. gelenkt wurde. Angefangen wurde dabei mit „Faktenfragen“, welche vor allem Fakten zum Vergleich und zur Einordnung der Institution bzw. der Nutzer erheben sollten. Diese hätten theoretisch bereits vor dem Interview beantwortet werden können, diese Möglichkeit wurde jedoch von keinem der Interviewten wahrgenommen, so dass diese am Anfang des Interviews standen. Anschließend folgten umfangreichere oder thematisch zusammengehörige

³³ Für eine Tabelle der gespiegelten Fragen siehe Abschnitt 9.8. Gespiegelte Fragen.

³⁴ Herzlichen Dank dazu an Robert Wein (Fotothek Stiftung Stadtmuseum Berlin) und Johanna Gentz (Masterstudentin Geschichtswissenschaft Humboldt-Universität zu Berlin) für die Rückmeldungen bei den Pre-Tests.

Fragen, die ausführlichere Antworten benötigten. Hinzu kamen immer wieder einige „Meinungsfragen“, in denen die Befragten nicht nur ihrem Umgang mit einem Thema (z.B. Open Access), sondern auch ihre Meinung dazu äußern sollten.

Die Fragen wurden jeweils gestellt und anschließend die Antworten ohne Zwischenfragen abgewartet, um auch Sichtweisen zum Thema, die zuvor vielleicht nicht bedacht worden waren, zu hören und die Interviewpartner nicht nur auf die Betrachtungsweisen, wie sie in dieser Arbeit dargestellt sind, festzulegen.³⁵ Erst anschließend wurden gegebenenfalls Nachfragen gestellt, die eine bestimmte Sichtweise des Themas konkretisierten. Nach- und Verständnisfragen der Interviewpartner wurden sofort beantwortet.

Leitfadeninterviews leiden oft unter dem Problem, dass die Fragen nur abgearbeitet werden, ohne konkret auf das Gesagte einzugehen (vgl. Werner 2013, S. 133f.). Um dem zu begegnen wurden, wenn thematisch passend, Zwischenfragen gestellt, die beschriebene Aspekte weiter vertieften.

Nach Abschluss der Transkriptionen wurden diese den jeweiligen Interviewpartnern, wenn von diesen gewünscht, zur Kontrolle zugesandt.

5.3. Angebote und Möglichkeiten von Kulturerbeeinrichtungen.

5.3.1. Digitale Sammlungen der untersuchten Institutionen.

Alle der in dieser Arbeit untersuchten Gedächtnisinstitutionen bieten ihre Digitalisate auch auf eigenen Online-Sammlungen an, welche daher kurz betrachtet werden sollen. Die Informationen, welche von den Websites gesammelt werden, sollen vor allem dem Vergleich der digitalen Angebote dienen. Eine ausführliche Evaluation der Angebote ist daher nicht nötig, auf klassische Evaluationsmetriken wie Effizienz, Effektivität oder Benutzbarkeit wird hier daher verzichtet.

Die digitalen Sammlungen stellen eine Art von Informationssystemen dar. „Informationssysteme werden weitgefasst als Systeme verstanden, die Informationsressourcen in Sammlungen ordnen, speichern, repräsentieren und für ihre Nutzer zugänglich machen“ (Petras 2013, S. 368). Ihre wichtigste Funktion ist die Zugänglichmachung von Informationen, wobei sich die Qualität durch die Inhalte und die Funktionalitäten bestimmt (vgl. Petras 2013, S. 368).

5.3.1.1. Methodik Webevaluation.

Die Evaluation von Informationssystemen untersucht die Qualität dieser nach unterschiedlichen Aspekten, wobei man grob zwischen Untersuchungen des Systems (system-centered), dem Verhalten der Nutzer (human-centered) und der Bedienbarkeit und Nützlichkeit der Funktionen und Eigenschaften (usability-centered) unterscheiden kann. Hinzu kommen anthropologische, soziologische oder ökonomische Ansätze (vgl. Petras 2013, S. 368f.; vgl. Saracevic 2004, S. 4f.). Je nach Ausrichtung können daher das System und seine Komponenten, das Interface und die Funktionen, die Nutzungsmöglichkeiten oder Nutzungskontexte zentral für die Evaluation sein (vgl. Fuhr et al. 2001, S. 191). Der Evaluationsprozess lässt sich dabei durch die Komponenten des

³⁵ Dies betraf vor allem die offenen Fragen oder Meinungsfragen, wie z.B. Frage N-4.2 „Was sind für Sie die Forschungsdaten der Geschichtswissenschaft?“

Systems, die Ziele des Evaluation sowie ihre Kriterien, die Methoden zur Datensammlung und den Kontext beschreiben (vgl. Petras 2013, S. 369).

Für diese Arbeit ist eine usability-zentrierte Evaluation angebracht, da die Inhalte und Funktionen der einzelnen digitalen Sammlungen betrachtet werden sollen. Das System von seiner technischen Seite her ist hierbei weniger wichtig, die Nutzer werden in den Interviews später befragt.

Zur Evaluation der digitalen Sammlungen soll hier das Evaluationsmodell von Fuhr et al. (vgl. Fuhr et al. 2001) als Grundlage genommen werden, welches die Daten bzw. Sammlung, die Technologie sowie Nutzer und Nutzung als Hauptdimensionen einer digitalen Sammlung ausmacht (vgl. Fuhr et al. 2001, S. 198). Da die Evaluation in dieser Arbeit jedoch nur vergleichende Fakten sammeln soll, werden nur einige Aspekte betrachtet. Andere vertiefende Schritte wie z.B. die unterschiedlichen Anforderungen und Erwartungen an die Nutzbarkeit eines Systems von Seiten der Nutzer und der Betreiber bzw. Designer entfallen. Zum Bereich der Daten bzw. Sammlung stellen sich daher die Fragen, wonach sich die Zusammenstellung der Objekte richtet (vgl. Fuhr et al. 2001, S. 195), welche bibliographischen Angaben gemacht sowie welche Metadaten angeboten werden (Fuhr et al. 2001, S. 193 sowie S. 195). Die Bereiche Technologie und Nutzung behandeln beide Aspekte zur Nutzung der Daten und dem Informationszugang und sind hierbei daher durch die Fragen, welche Bearbeitungsfunktionen zur Verfügung stehen (vgl. Fuhr et al. 2001, S. 195) und welche Sucheinstiege sich dem Nutzer bieten (vgl. Fuhr et al. 2001, S. 193f. sowie S. 196) vertreten. Andere Aspekte wie die verwendeten Protokolle bleiben hierbei unberücksichtigt, Fragen nach den Nutzergruppen werden in den Interviews ausführlicher gestellt. Hinzu kommen jedoch noch Aspekte der DFG-Mindestanforderungen an die Bereitstellungssysteme von Digitalisaten, welche in den Fragen zu Schnittstellen (vgl. Fuhr et al. 2013, S. 193 und S. 196), der Download- und Druckfunktion, einer Ansicht im DFG-Viewer und einer Feedback-Möglichkeit berücksichtigt sind (vgl. DFG 2016, S. 44f.).³⁶

5.3.1.2. Auswertung Webevaluation.

Bei der Sammlung Online der Stiftung Stadtmuseum Berlin³⁷ werden die Digitalisate des Stadtmuseums Berlin angeboten. Dabei wird ein Überblick über die Sammlungsbereiche des Museums gegeben, wobei einige Sammlungen stärker vertreten sind als andere. Einige Objekte sind zudem in Gruppen für ein einfacheres Browsing (nach Menschen, Orten, Epochen oder ihren Digitalisierungsprojekten) sortiert. Zu jedem Objekt werden bibliographische Angaben zum Künstler, Titel, Erscheinungsjahr und -ort, Material, Technik, Objektart und den Maßen gegeben, bei einigen Objekten sind auch wissenschaftliche Kommentare geschrieben.

Bearbeitungsfunktionen sind allgemein nicht möglich, je nach Medientyp können hochauflösende Bilder vergrößert, einige Objekte auch in einer 360°-Ansicht gedreht oder geblättert werden. Als Sucheinstieg bietet sich neben der einfachen Suche eine erweiterte

³⁶ Siehe 9.4. Evaluationsfragen zur Evaluation der Websites.

³⁷ Stadtmuseum Berlin. Sammlung Online, <https://sammlung-online.stadtmuseum.de/> (abgerufen am 24.05.2019, 13:45).

Suchfunktion, welche die Parameter Sammlung, Titel, Material, Technik, Datierung, Personen, Objektart und Schlagwort zur Verfügung stellt. Zum Browsen können die schon beschriebenen Projektkategorien Menschen (z.B. Berliner Typen, Bürger, Persönlichkeiten, ...), Orte (Stadtansichten, Verkehr, Rathaus, ...), Epochen oder die Digitalisierungsprojekte gewählt werden, zudem wird ein Best-Of der Bestände angeboten.³⁸

Die Funktionen der DFG-Mindestanforderungen an Bereitstellungssysteme werden kaum erfüllt: Die Daten werden über keine Schnittstelle angeboten (nur ein manuelles Web Mining über den Seitencode ist möglich), eine Ansicht im DFG-Viewer ist nicht möglich und es gibt weder eine Downloadfunktion noch eine eigene Druckfunktion, wobei im Versuch die Druckfunktion des Browsers (Mozilla Firefox 68.0.1 (64-Bit) auf Microsoft Windows 7 (64 Bit)) gute Ergebnisse für die Objektseiten lieferte. Einzig eine Feedback-Möglichkeit ist bei allen Objekten gegeben, zudem sind Kontaktinformationen dazu im Impressum vorhanden.

Die Digitalisierten Sammlungen der Staatsbibliothek zu Berlin, Preußischer Kulturbesitz³⁹ bieten historische Drucke, Handschriften und Nachlässe aus den Digitalisierungsprojekten der Staatsbibliothek an. Ein Schwerpunkt liegt auf deutschsprachigen Drucken des 16. bis 19. Jahrhunderts aus allen Fachrichtungen, hinzu kommen Ostasiatica, abendländische Handschriften und Musikhandschriften.⁴⁰ Bibliothekstypisch werden umfangreiche bibliographische Angaben gemacht, beginnend bei Titel, Autor und Erscheinungsjahr über Standort und Schlagworte bis hin zum Strukturtyp und der Anzahl der gescannten Seiten.⁴¹ Auch Angaben zum Digitalisierungsprojekt oder dem Indexierungsdatum des Digitalisates finden sich. Wenn die Lizenz vom Standard (Public Domain) abweicht, so wird dies hier gekennzeichnet. Alle Metadaten lassen sich dabei auch als METS (XML) oder IIIF Manifest (JSON) herunterladen.

Als Bearbeitungsfunktionen stehen Vergrößern, Drehen sowie einfache Bildmanipulation (Helligkeit, Kontrast, Sättigung, Farbinvertierung) zur Verfügung. Die Strukturdaten lassen sich anzeigen und zur Navigation verwenden. Als Sucheinstieg wird ein einfacher Suchschlitz geboten, wobei sich die durchsuchten Sammlungen noch eingrenzen lassen. Die Ergebnisliste daraus lässt sich anschließend durch Facetten zu Strukturtypen und Treffern mit Volltext sowie nochmals einer Sammlungsliste eingrenzen. Allerdings wird auch im regulären Bibliothekskatalog (StaBiKat)⁴² auf Digitalisate von Werken in der Sammlung hingewiesen. Hier stehen dann die vollen Suchmöglichkeiten einer erweiterten Suche im klassischen OPAC oder als RDS (stabikat+) zur Verfügung, es wird aber im gesamten Bestand, nicht nur der Digitalisierten Sammlung gesucht. Die Suche lässt sich jedoch auf E-Books/Online-Ressourcen eingrenzen. Ansonsten bietet sich eine einfache Browsingfunktion, die aus der Auswahl einer der Suchkategorien besteht.

³⁸ Für eine komplette Auflistung aller Kategorien siehe Abschnitt 9.5.1. Evaluation Stiftung Stadtmuseum Berlin, Sammlung Online.

³⁹ Staatsbibliothek zu Berlin. Digitalisierte Sammlungen, <https://digital.staatsbibliothek-berlin.de/> (abgerufen am 20.06.2019, 23:34).

⁴⁰ Vgl. Staatsbibliothek zu Berlin. Über digitalisierte Sammlungen, <https://digital.staatsbibliothek-berlin.de/ueber-digitalisierte-sammlungen> (abgerufen am 20.06.2019, 23:36).

⁴¹ Für eine komplette Auflistung aller erfassten Metadaten siehe Abschnitt 9.5.2. Evaluation Staatsbibliothek zu Berlin, Preußischer Kulturbesitz, Digitalisierte Sammlungen.

⁴² Staatsbibliothek zu Berlin. StaBiKat, <http://stabikat.de/> (abgerufen am 20.06.2019, 23:45).

Wie beschrieben lassen sich die Metadatensätze als METS-XML oder IIIF exportieren. Eine OAI-PMH-Schnittstelle wird nicht in den digitalen Sammlungen angezeigt, ist jedoch verfügbar.⁴³ Auch eine Downloadfunktion besteht, wobei entweder Einzelseiten je als TIFF oder JPEG oder das gesamte Buch auch als TIFF oder JPEG in einer ZIP-Datei oder als PDF zur Verfügung stehen. Wenn vorhanden kann auch der Volltext aus OCR-Verfahren als Datei (ALTO-XML) heruntergeladen werden. Eine Druckfunktion besteht nicht und auch die Druckfunktion des Browsers funktioniert nicht gut, jedoch lassen sich die entsprechenden Daten auch als PDF laden und dann ausdrucken. Neben der Ansicht in den Digitalisierten Sammlungen stehen auch der DFG-Viewer und der Mirador-Viewer zur Verfügung, eine Feedback-Möglichkeit besteht nur für das gesamte Angebot, nicht für einzelne Objekte.

In der Digitalen Sammlung der Universitätsbibliothek der Humboldt-Universität zu Berlin⁴⁴ werden die Digitalisate aus den historischen Sammlungen der Universitätsbibliothek präsentiert. Auch hier sind die bibliothekstypischen bibliographischen Angaben gegeben, also Titel, Autor, Erscheinungsjahr, -ort und Verlag, zudem weitere Angaben wie Strukturtyp, Identifier und ergänzende Informationen.⁴⁵ Weitere Metadaten wie zusätzliche Beschreibungen, Titelverknüpfungen oder Angaben zu Sprache, Seitenzahl oder Format finden sich im OPAC, in dem man sich die Objekte auch anzeigen lassen kann.

Die Bearbeitungsfunktionen begrenzen sich auf eine stufenlose Vergrößerung und das Drehen der Anzeige um je 90°. Der Sucheinstieg geschieht vor allem über Browsing, wobei die Kategorien Sammlungen (z.B. Gelehrtenbibliotheken oder historische Drucke), Wissensgebiete (also die wissenschaftlichen Disziplinen), Projekte und neu digitalisierte Bestände zur Auswahl stehen. Für die klassische Suche steht ein einfacher Suchschlitz zur Verfügung, über Facetten können dann Dateitypen, Strukturtypen, beteiligte Personen sowie das Entstehungsjahr gefiltert werden. Alternativ ist eine umfangreiche Suche über den regulären OPAC/RDS der Universitätsbibliothek möglich.

Als Schnittstelle wird OAI-PMH angeboten, wobei hier die Formate MARC XML, Dublin Core XML oder ESE (Europeana Semantic Elements, der Vorgänger des aktuellen Europeana Data Model EDM) zur Verfügung stehen. Zudem lassen sich die Metadaten als METS oder das gesamte Werk als PDF downloaden. Eine eigene Druckfunktion ist nicht vorhanden und auch die Druckfunktion des Browsers wird nicht optimal bedient, ein Druck der Daten ist dennoch aus dem PDF heraus möglich. Eine Ansicht der Objekte im DFG-Viewer wird unterstützt, die Möglichkeit zum Feedback ist gegeben.

Alle der hier untersuchten Websites bieten als digitale Sammlungen die Digitalisate ihrer jeweiligen Einrichtungen an. Ein deutlicher Unterschied zeigt sich dabei zwischen den Angeboten der beiden untersuchten Bibliotheken und des Stadtmuseums. Während das Museum nur grundlegende bibliographische Metadaten angibt, wie es in Inventarbüchern

⁴³ Siehe Abschnitt 5.3.3.5. Interoperabilität und Schnittstellen, Portale (I-5.) sowie 9.9.2. Interview Staatsbibliothek zu Berlin, Preußischer Kulturbesitz..

⁴⁴ Universitätsbibliothek der HU Berlin. Digitale Sammlungen, <https://www.digi-hub.de/viewer/> (abgerufen am 14.06.2019, 12:30).

⁴⁵ Für eine komplette Auflistung aller erfassten Metadaten siehe Abschnitt 9.5.3. Evaluation Universitätsbibliothek der Humboldt-Universität zu Berlin, Digitale Sammlungen (digi-hub).

üblich ist, sind diese in den Bibliotheken weit ausgeprägter und entsprechen den Angaben von Bibliothekskatalogen.

Diese Unterschiede zeigen sich auch bei den Funktionalitäten. Während die Angebote der Bibliotheken klar auf Bücher ausgelegt sind und daher die Möglichkeit des Blätterns, eine Anzeige von Strukturdaten und teilweise die Volltexte vorhanden sind, ist die Sammlung Online des Stadtmuseums eher auf Museumsobjekte ausgelegt. Bücher lassen sich zwar auch über eine einfache Blätterfunktion anschauen, es gibt aber auch die Möglichkeit, einige Objekte von allen Seiten zu betrachten. Einzig die Funktion zum Vergrößern der Ansicht ist bei allen Institutionen als grundlegendes Arbeitsmittel gegeben. Die Staatsbibliothek stellt hierbei zudem noch Funktionen der Bildmanipulation zur Verfügung. Die Suchfunktionalität ist interessanterweise bei allen drei Angeboten relativ ähnlich, es gibt die einfache Suche sowie eine erweiterte Suche in den jeweils vorhandenen bibliographischen Metadaten, wobei die bibliothekarischen Angebote hierbei auf den OPAC zurückgreifen. Ansonsten bieten alle Institutionen die Möglichkeit des Browsens, wobei diese bei der Universitätsbibliothek und dem Stadtmuseum ausgeprägter sind, die beiden Bibliotheken bieten zudem umfangreiche Facettierungsmöglichkeiten an.

Die Exportmöglichkeiten sind beim Stadtmuseums nicht gegeben bzw. nur durch Web Mining zu erreichen, während die Bibliotheken sowohl die Digitalisate wie die Metadaten in den für Bibliotheken typischen Formaten sowie über die aus dem Bibliotheksbereich stammende OAI-PMH-Schnittstelle zur Verfügung stellen. Dagegen ist die Feedback-Funktion im Museum stärker ausgeprägt und soll den Nutzern hier auch die Möglichkeit geben, Wissen über das spezifische Objekt an das Museum heranzutragen, eine Funktion, die im bibliothekarischen Bereich weniger Sinn machen würde.

Als einzige Einrichtung bietet die Staatsbibliothek zudem eine umfangreiche Anleitung für die Funktionen ihrer Digitalisierten Sammlungen an.

Insgesamt zeigt sich an den digitalen Angeboten, dass im Museum eher die Objekte im Vordergrund stehen (verschiedene Ansichtsmöglichkeiten, Feedback-Möglichkeit, ...), während die Bibliotheken den Schwerpunkt auf den Inhalt legen (Strukturdaten, Volltextdaten, ...).

5.3.2. Beschreibung der Fragen und des Aufbaus (Methodik) - Institutionen.

Dem Forschungsdatenlebenszyklus folgend beginnen die Interviews mit den Mitarbeitern von Institutionen mit der Auswahl der zu digitalisierenden Objekte.⁴⁶ Da der Auswahlprozess im Interview je genauer beleuchtet werden soll, gibt es hierzu keine einfachen Faktenfragen.

Die Kernfrage zum Aspekt der Auswahl von Objekten beschreibt den Auswahlprozess von Objekten für Digitalisierungsprojekte und in wie fern Kuratoren/Sammlungsleiter, Konservatoren, Mittelgeber und Nutzer berücksichtigt werden (I-1.1.). Dadurch sollen der Auswahlprozess sowie die beteiligten Akteure in Erfahrung gebracht werden. Dies ist für mehrere Bereiche interessant: Da in der Geschichtswissenschaft wie beschrieben⁴⁷ alles eine Quelle sein kann, könnten auch alle Objekte interessant für die Nutzer sein. Hinzu

⁴⁶ Obwohl während der Interviews wie beschrieben erst die Faktenfragen gestellt wurden und erst dann die weiteren Fragen kamen, sollen die Fragen bei der Begründung und Auswertung zur besseren Übersichtlichkeit thematisch, also gemeinsam, behandelt werden.

⁴⁷ Siehe Abschnitt 2.2. Forschungsdaten in der Geschichtswissenschaft.

kommt der Aspekt der Retrodigitalisierung als Mittel der Bestandserhaltung für gefährdete Bestände⁴⁸ sowie eigene Bedürfnisse und die Anforderungen von Mittelgebern, z.B. durch bewilligte Digitalisierungsprojekte, welche alle Einfluss auf Auswahlentscheidungen haben könnten.

Durch die zweite Frage (I-1.2.), welche in dieser Form auch auf Nutzerseite gestellt wird, soll geklärt werden, wie die Institutionen mit Dubletten umgehen, ob beide Exemplare oder nur eines digitalisiert wird oder ein inhaltlich beispielhaftes Objekt aus mehreren Dubletten erstellt wird. Gerade vor dem Hintergrund der historischen Werke, welche sich meist allein schon auf Grund ihrer Provenienz unterscheiden (vgl. Stauffer 2016), und ihrer Verwendung in der Wissenschaft ist dies ein wichtiger Aspekt, der beschreibt, ob bei der Auswahl der Inhalt oder das Objekt selber im Fokus steht.

Darauf folgt der Fragenabschnitt zum eigentlichen Digitalisierungsvorgang, welcher eine Reihe von Faktenfragen beinhaltet. Diese fragen nach angewandten externen Standards wie z.B. den DFG-Praxisregeln (If-2.1), den konkreten Anforderungen an die Bilddigitalisierung (If-2.2.), ein Aspekt, der auch auf der Nutzerseite abgefragt wird, der Anzahl der insgesamt angebotenen Digitalisate (If-2.3.) sowie der jährlich neu digitalisierten Bestände (If-2.4.) und organisatorischen Aspekten wie dem Vorhandensein von Born Digitals in den digitalen Angeboten (If-2.5). Hinzu kommt die Frage, ob die Digitalisierung im Haus oder durch Dienstleister durchgeführt wird (If-2.6.) und, wenn sie im Haus stattfindet, wie viele Mitarbeiter daran beteiligt sind. Diese Fragen sollen ein Bild vom Digitalisierungsvorgang in der jeweiligen Einrichtung geben, um sie besser einordnen zu können. Frage If-2.7. soll klären, ob es Maßnahmen gibt, um Doppeldigitalisierungen zu vermeiden, da der Digitalisierungsvorgang immer einen finanziellen und zeitlichen Kostenfaktor sowie eine Belastung der Originale darstellt. Diese Maßnahmen können auch Absprachen mit anderen Einrichtungen beinhalten, wie es vom DBV oder der KII gewünscht wird (vgl. DBV 2018, S. 19; vgl. KII 2011, S. 37).

In den weiteren Fragen zur Digitalisierung soll geklärt werden, ob die Digitalisierung immer vom Original oder, wenn vorhanden, auch vom Film/Microfiche erfolgt (I-2.1.). Die Umwandlung von analog auf digital führt immer zu Qualitätsverlusten, durch den Zwischenschritt über einen Film wird dieser, besonders bei Farbdigitalisaten, noch verstärkt (vgl. Czmiel et al. 2005, S. 24; Lang/Bohne-Lang 2019, S. 150). Allerdings kann vom Film in der Regel schneller digitalisiert werden. Diese Frage schließt somit an vorherige Fragen zum Verhältnis Objekt/Inhalt an. Die nächste Frage (I-2.2.) konkretisiert die Rolle der DFG-Praxisregeln und deren Relevanz für die Einrichtung, da diese als Standard bei DFG-geförderten Projekten einen Quasi-Standard bilden, wobei eine Orientierung an diesem eine starke Vereinheitlichung der Digitalisierungsergebnisse hätte. Wichtig für die Digitalisierungsergebnisse ist auch die Qualitätssicherung, die auch ein wichtiger Teil des Forschungsdatenmanagements ist. Wenn Digitalisate als Forschungsdaten behandelt werden sollen, so müssen dazu auch Qualitätssicherungsmaßnahmen getroffen werden. Die Frage I-2.3. beschäftigt sich mit diesen, sowohl im Workflow als auch auf technischer Ebene durch die Verwendung von Testtafeln für Farbigkeit und Schärfe, Farbmesssysteme (ColorChecker) und ICC-

⁴⁸ Siehe Abschnitt 3. Digitalisate als Forschungsdaten.

Farbprofile (vgl. Lang/Bohne-Lang 2019, S. 161). Wichtig für die Nachnutzung von Forschungsdaten ist zudem die Persistenz, weshalb Digitalisate im Nachhinein nicht verändert werden sollten bzw. bei Veränderungen eine Versionshistorie erfasst werden muss. Frage I-2.4. klärt das Verhalten der Institutionen hierbei.

Damit leitet die Frage auch zum nächsten Abschnitt, der Dokumentation und den Metadaten der Digitalisate über. Die Faktenfragen klären hierbei zuerst, welche Metadaten überhaupt Verwendung finden (If-3.1.). Dies ist zentral für die Erfassung und Nutzung der Digitalisate und wird daher gespiegelt auch auf der Nutzerseite gefragt. Des weiteren wird geklärt, ob alle erfassten Metadaten den Nutzern frei zur Verfügung stehen (If-3.2.) und nach welchen Metadatenschemata und Erschließungsstandards die Daten beschrieben werden (If-3.3. und If-3.4.), wobei zum Vergleich auch das favorisierte Metadatenformat der Nutzer abgefragt wird. Auch die Erschließung des Volltextes und sein Format werden, sowohl auf Seiten der Institution wie der Nutzer, thematisiert (If-3.5.). In den weiteren Fragen werden diese Aspekte dann spezifiziert, indem nach den gegebenen Informationen zum Kontext der Digitalisierung und der Dokumentation des Prozesses gefragt wird (I-3.1.). Besonders die Dokumentation des Entstehungsprozesses und der Provenienz sind wichtig, um einen Kontext und Transparenz zu schaffen, was nicht nur das Vertrauen in das Digitalisat und seine Authentizität stärkt (vgl. Schubert 2018, S. 242), sondern auch von der DFG gefordert wird (vgl. DFG 2016, S. 7), weshalb sie in gespiegelter Form auch den Nutzern gestellt wird. Frage I-3.2. stellt die erste Meinungsfrage da, in welcher die Interviewten ihren Standpunkt erläutern sollen, ob umfangreiche Metadaten bei Digitalisaten einen Ausgleich zum Verlust der analogtypischen Besonderheiten wie z.B. Materialität schaffen können. Diese Besonderheiten sind Untersuchungsgegenstand einiger Forschungsdisziplinen wie z.B. der Buchwissenschaft, hier soll daher auch an Metadaten gedacht werden, die möglicherweise zukünftig zu vergeben sind, um noch mehr Daten zur Verfügung zu stellen. Auch die Meinung zu der Möglichkeit, ob ein Digitalisat ein analoges Original ersetzen kann (vgl. Schubert 2018, S. 240) soll kurz erläutert werden. Diese Frage wird in der Form auch den Nutzern gestellt, um die Standpunkte anschließend abgleichen zu können.

Nach Herstellung und Erschließung ist die Veröffentlichung der nächste Abschnitt des Forschungsdatenlebenszyklus. Hierzu wird nur eine Faktenfrage gestellt, nämlich nach den in den digitalen Angeboten verwendeten Lizenzen und welche Lizenz hier aus welchem Grund favorisiert wird (If-4.1.).

Etwas genauer wird in diesem Bereich gefragt, in welcher Form die Daten veröffentlicht werden. Dabei soll geklärt werden, ob alle digitalisierten Objekte (soweit rechtlich möglich) auch in den digitalen Angeboten (frei) zur Verfügung stehen (I-4.1.) und ob diese dann auch in der höchstmöglichen Qualität (Master) angeboten werden (I-4.2.). Wenn dies nicht der Fall ist, so soll kurz erklärt werden, worin sich die angebotenen Digitalisate unterscheiden und was nötig ist, um die Master-Dateien zu erhalten. So soll herausgefunden werden, ob es technische Einschränkungen bei den angebotenen Digitalisaten gibt. Außerdem wird nach der Rolle von Zertifikaten für

Forschungsdatenrepositorien wie z.B. dem Data Seal of Approval oder dem DINI-Zertifikat gefragt (I-4.3.), welche teilweise als Mindestanforderungen an Forschungsdatenrepositorien beschrieben werden (vgl. Pempe 2012, S. 147). Wenn Digitalisate als Forschungsdaten betrachtet werden sollen, so stellen die digitalen Sammlungen damit eine Art Forschungsdatenrepositorien dar, so dass hier ermittelt wird, ob die digitalen Angebote sich selber als Repositorien verstehen und daher nach deren Qualitätszertifikaten streben.

Auch die nächste Fragengruppe steht im Zusammenhang mit der Veröffentlichung, indem Aspekte zur Interoperabilität, Schnittstellen und Portalen abgefragt werden. Die Faktenfragen klären dabei, ob bereits maschinenlesbare Schnittstellen zum Harvesting unterstützt werden (If-5.1.), eine Anforderung, die durch die Verbreitung der auch quantitativ arbeitenden Digital Humanities zunehmend auch in der sonst eher qualitativ arbeitenden Geschichtswissenschaft vorkommt, und ob die Digitalisate auch auf andere Plattformen als die eigene digitale Sammlung hochgeladen werden (If-5.2.), was Sichtbarkeit, Auffindbarkeit und den Sucheinstieg verbessern könnte. Da den Nutzern die Fragen nach ihren favorisierten Portalen und ihrer Nutzung von Harvesting gestellt werden, kann so abgeglichen werden, ob sich die von Institutionen und Nutzern bevorzugten Portale und Schnittstellen decken.

Vertiefend wird nach (internen wie externen) Verlinkungen zwischen den Objekte gefragt (I-5.1.), welche noch leichter Zusammenhänge und Kontextinformationen zur Auswertung von Objekten liefern würden. Die Bedeutung von Verlinkungen wird daher auch bei den Nutzern abgefragt, um hier einen Abgleich zu schaffen. Des weiteren wird thematisiert, ob die jeweilige Einrichtung ihre Angebote auch auf weniger wissenschaftlichen, aber möglicherweise populäreren Plattformen wie Wikipedia anbietet (I-5.2.), welche im Alltag eher genutzt werden als spezialisierte Dienste (vgl. KII 2011, S. 151). Dies ist insofern interessant, da Informationen in den sonst bespielten Angebote wie Europeana vor allem durch Gedächtnisinstitutionen bzw. deren fachlich qualifizierten Mitarbeiter erstellt werden, während bspw. Wikipedia allen Nutzern die Möglichkeit der Inhaltserstellung und Mitarbeit bietet.

Den nächsten Schritt im Forschungsdatenlebenslauf stellt die Speicherung, Pflege und Archivierung der Daten dar. Da diese Fragen miteinander zusammenhängen und keine einfachen Antworten ermöglichen, werden hierzu keine Faktenfragen gestellt. Um in Erfahrung zu bringen, wie die Langzeitarchivierung in der jeweiligen Einrichtung aussieht, wird statt dessen nach speziellen Strategien zur Langzeitarchivierung gefragt (I-6.1.) und wie die Backup-Strategie gestaltet ist (I-6.2.), was auch Abstände zwischen den Sicherungen, Redundanzen und mögliche spezielle Formate beinhaltet. Daran anschließend folgen zwei Meinungsfragen, in denen die Interviewten erklären sollen, ob es sich ihrer Meinung nach im Bezug auf Aufwand, Finanzen und Verwaltung lohnt, Digitalisate überhaupt langfristig zu archivieren oder diese bei Bedarf neu vom Original aufnehmen zu lassen (I-6.4.). Zudem wird gefragt, ob qualitativ bessere Neuaufnahmen bei Verfügbarkeit an die Stelle der alten Digitalisate treten sollten oder ob beide Versionen, wie es bei Forschungsdaten allein schon auf Grund der Referenzierbarkeit

üblich ist, erhalten werden sollten (I-6.5.). Abschließend wird der Umgang mit veralteten Digitalisaten thematisiert (I-6.3.), wobei veraltet hier nicht unbedingt eine Verhinderung der Nutzbarkeit darstellen muss, sondern die Daten einfach nur nicht mehr aktuellen Nutzererwartungen entsprechen (vgl. Ludwig/Enke 2013, S. 36).

Die letzte Fragengruppe beschäftigt sich mit organisatorischen Fragen der Digitalisierung. Die Faktenfragen drehen sich dabei um die unterstützenden Fördereinrichtungen (If-7.1.), welche mit ihren Profilen und Zielen auf die Digitalisierungsstrategie der Einrichtung einwirken könnten, des weiteren um die Möglichkeit, neue Funktionen im System der digitalen Sammlungen zu integrieren (If-7.2.), wodurch auf sich ändernde Anforderungen der Nutzer reagiert werden kann, und um Erhebungen, welche Nutzergruppen die digitalen Angebote (am meisten) nutzen (If-7.3.), da sich je nach Nutzergruppe die Anforderungen unterscheiden.

Für Forschungsdaten wird oftmals ein Datenmanagementplan gefordert, so dass dieser auch für Digitalisate vorliegen sollte, wenn diese als Forschungsdaten zu verstehen sind. Daher wird nach dem Vorliegen eines Datenmanagementplans gefragt und, wenn einer vorhanden ist, welche Aspekte er abdeckt und wer für die Umsetzung verantwortlich ist (I-7.1.). Zudem wird nach der Zielgruppe gefragt, an der sich die Einrichtung am meisten orientiert (I-7.2.) und ob diese Nutzergruppen bei Entscheidungen im Digitalisierungsprozess, z.B. bei der Auswahl der als nächstes zu digitalisierenden Objekte, mit einbezogen werden (I-7.3.), wie es teilweise in der Literatur gefordert wird (vgl. Ludwig/Enke 2013, S. 27). Zum Vergleich wird diese Frage auch den Nutzer gestellt. Abschließend folgen drei Meinungsfragen, wovon die erste auch den Nutzern gestellt wird, nämlich wie man zu dem Vorschlag steht, Gedächtniseinrichtungen sollten Forschungsdaten, die aus ihren Digitalisaten hervorgehen (z.B. Korpora, Textauswertungen, ...) selber sichern und verfügbar machen und so die Rolle eines Repositoriums übernehmen, damit diese auch als Forschungsdaten weitergenutzt werden können (I-7.4.). Damit könnte dem Problem begegnet werden, dass Daten aus der oft projektgetriebenen Forschung nach Projektende verloren gehen, da sie hier aufbewahrt werden würden (vgl. Sahle/Kronenwett 2013, S. 83). Man würde jedoch in eine gewisse Konkurrenz zu disziplinären Repositorien oder den Repositorien der Forschungsinstitutionen treten. In der nächste Meinungsfrage sollen die Interviewten als leitende Mitarbeiter der Digitalisierungsstellen ihrer Einrichtungen beschreiben, was sich ihrer Meinung nach bei der Digitalisierung in Gedächtniseinrichtungen verbessern sollte (I-7.5.) und abschließend darlegen, welche Möglichkeiten, z.B. im Bezug auf Digitalisierungsgeschwindigkeit, Erschießungstiefe, Funktionen etc., ihre Einrichtung bei unbegrenzten Mitteln hätte (I-7.6.).

5.3.3. Auswertung der Interviews - Institutionen.

5.3.3.1. Auswahl (I-1.).

Der Auswahlprozess sieht in allen Einrichtungen relativ ähnlich aus, auch wenn sich bestimmte Spezifika feststellen lassen. Grundsätzlich werden in Absprache mit den jeweiligen Sammlungen Objekte zu Digitalisierungsprojekten zusammengefasst, nach Kriterien wie Unikalität, Fragilität und aktueller oder wissenschaftlicher Relevanz

ausgewählt und einer konservatorischen Prüfung unterzogen. Gewisse Unterschiede gibt es bei weiteren, pragmatischen Kriterien: Während die Stiftung Stadtmuseum Berlin (SSB) und die Universitätsbibliothek (UB⁴⁹) für große Digitalisierungsprojekte auf finanzielle Fördermittel angewiesen sind und die finale Auswahl der Projekte daher durch die Mittelgeber geschieht bzw. die Projekte zu den entsprechenden Ausschreibungen geschrieben werden, ist die Staatsbibliothek zu Berlin (SBB) durch ihr eigenes großes Digitalisierungszentrum unabhängiger. Finanzielle Mittel spielen hier zwar auch eine Rolle, jedoch können auch Projekte aus eigenen Mitteln realisiert werden. Es muss allerdings darauf geachtet werden, dass nicht nur wie in den anderen Einrichtungen genügend Kapazitäten zur Verarbeitung der Digitalisate vorhanden sind, sondern auch genügend Kapazitäten im Digitalisierungszentrum (Scanner, Mitarbeiter, ...). Hinzu kommen in allen Einrichtungen Sicherungsdigitalisierungen zur Bestandserhaltung und meist kostenpflichtige Digitalisierungen im Nutzauftrag.

Auch der Umgang mit Dubletten ist in allen untersuchten Einrichtungen einheitlich, grundsätzlich würde man immer das vollständigste oder am besten erhaltene Exemplar digitalisieren. Ausnahmen können, besonders in den historischen Sammlungen, Exemplarspezifika wie handschriftliche Annotationen oder, eher im museologischen Bereich, unterschiedliche Materialien sein, wobei hier eher von mehreren Exemplaren als wirklichen Dubletten zu sprechen ist.

Das Erstellen eines Digitalisates aus unterschiedlichen Exemplaren, also das Erzeugen eines „schöneren Exemplars [, als im] Buchbestand vorhanden“ (Müller, SBB, I-1.2.), wird überall abgelehnt, wenn es sich nicht um Ausnahmen wie das Erstellen eines gesamten, sonst nicht mehr existenten Zeitschriftenbestandes aus unterschiedlichen Sammlungen handelt.

5.3.3.2. Digitalisierung (I-2.).

Als Mindeststandards für die Digitalisierung werden in allen Einrichtungen die DFG-Richtlinien beschrieben, also konkret eine TIFF-Datei mit 300 dpi, 24 Bit Farbtiefe im RGB-Farbraum. Überall stellt dies jedoch nur die Untergrenze dar, objekt- oder projektbezogen können auch höhere Standards verwendet werden. Die Vorteile der DFG-Richtlinien werden in einheitlichen Mindeststandards, einer Vergleichbarkeit mit anderen Einrichtungen, Interoperabilität und Sicherheit gesehen, da die Daten so langfristig präsentiert werden können, z.B. im DFG-Viewer. Die SBB berücksichtigt zunehmend auch weitere, technische Standards wie die Metamorfoze-Guidelines oder den ISO 19264-1 Standard, da die DFG-Richtlinien kein technisches Mindestmaß für die Abbildungsqualität vorgeben.

Unterschiede zwischen den Einrichtungen gibt es dagegen bei der Größe ihrer Digitalisierungseinrichtungen und der digitalen Sammlungen.

Mit 154.159 digitalisierten Werken stellt die SBB die größte der untersuchten Einrichtungen dar (Stand: 19.06.2019). Gezählt werden dabei nur die Werke, in einzelnen

⁴⁹ Die sonst übliche Abkürzung der Universitätsbibliothek der Humboldt-Universität ist UB der HU. Auf Grund der Übersichtlichkeit im Fließtext wird in dieser Arbeit jedoch die verkürzte Form UB verwendet, welche sich in diesem Rahmen jedoch explizit auf die Universitätsbibliothek der Humboldt-Universität bezieht.

Bildern bestehen ihre Digitalisierten Sammlungen aus ca. 20 Mio. Dateien. Pro Jahr werden ca. 10.000-15.000 Vorgänge neu digitalisiert, was anderthalb Mio., bei viel Digitalisierung vom Rollfilm auch zwei Mio., Dateien entspricht. Die digitalisierten Sammlungen enthalten dabei nur Datensätze, die im Digitalisierungszentrum erzeugt wurden, Born Digitals befinden sich nicht in dieser Sammlung, da diese als Bibliotheksbestand im Katalog zu finden sind. Die Digitalisierung findet wenn möglich im eigenen Digitalisierungszentrum statt, da dies bestandsschonender ist. Dienstleister werden nur für die wenigen Fälle in Anspruch genommen, bei denen es am technischen Equipment fehlt. Im Digitalisierungszentrum arbeiten immer 20 bis 25 Mitarbeiter, im gesamten Haus sind am Digitalisierungsprozess über 100 Personen beteiligt.

Das Stadtmuseum hat 60.610 Objekte in der Sammlung Online (Stand: 20.05.2019), von denen die meisten ein, einige auch bis zu drei Bilder haben. Pro Jahr werden ca. 18.500 Objekte neu digitalisiert, Born Digitals sind nur in Teilbereichen der Sammlungen (Theaterfotografie) oder als dokumentarische Quellen vorhanden. Die Digitalisierungen werden entweder ausgeschrieben (besonders Massendigitalisierungen) oder finden intern statt, wobei direkt drei Mitarbeiter mit der Erstellung und Erfassung der Digitalisate beschäftigt sind (Scanoperator, Fotograf, Datenerfasser). Am gesamten Digitalisierungsprozess sind um die 40 Mitarbeiter beteiligt.

In der digitalen Sammlung der UB sind ca. 7.000 Werke online (Stand:12.06.2019), welche jeweils aus einer unterschiedlichen Zahl an Bildern bestehen. Pro Jahr kommen ca. 1.000 neu digitalisierte Werke dazu, wobei der eigentliche Digitalisierungsvorgang bereits länger zurück liegen kann, die Publikation in der digitalen Sammlung aber erst das Ende des Digitalisierungsvorgangs darstellt. Born Digitals kommen kaum vor, nur im Bereich der Professorenporträts sind einige vorhanden. Die Digitalisierung findet sowohl im Haus als auch durch Dienstleister statt, wobei intern drei, inklusive der EDV viereinhalb, Personenäquivalente, verteilt auf mehrere Mitarbeiter, mit der Digitalisierung beschäftigt sind.

Entsprechend den Empfehlungen des DBV und der KII finden Kontrollen, ob zur Digitalisierung vorgesehene Objekte bereits woanders digital vorliegen, von den hier untersuchten Einrichtungen nur bei den Bibliotheken statt. Die Unikalität ist ein wichtiges Auswahlkriterium bei der SBB und schließt damit schon von sich aus Doppeldigitalisierungen aus, bei DoD-Aufträgen findet eine Kontrolle statt. An der UB ist eine solche Überprüfung Teil des Digitalisierungsworkflows. Da die Bücher jedoch auch historische Objekte sind, kann es auch Ausnahmen geben, z.B. wenn die Digitalisate in anderen Einrichtungen nicht den eigenen Qualitätsanforderungen entsprechen oder das Buch in einen bestimmten Provenienzzusammenhang gehört. Ziel im Museum ist es dagegen, die eigene Sammlung digital zur Verfügung zu stellen, so dass das Vorhandensein in anderen Einrichtungen kein Ausschlusskriterium ist. Kontrollen gibt es höchstens in einigen Projekten oder Sammlungsbereichen wie der Hausbibliothek.

Die Digitalisierung findet dabei in allen Einrichtungen grundsätzlich im Original statt, da die Qualität so besser ist und zusätzliche Informationen aufgenommen werden können, die bei alten Aufnahmen auf Film vielleicht fehlen. Ausnahmen bilden Fotos (SSB), da Abzüge hier bereits verlustbehaftete Derivate sind und (historische) Zeitungen (SBB), da

es hier eher auf den Textinhalt ankommt und diese in der Regel bereits verfilmt sind, eine Digitalisierung vom Rollfilm daher um einiges schneller geht. Zudem sind Zeitschriften oft stark in den Falz eingebunden und haben fragiles Papier, so dass konservatorische Gründe gegen das Original sprechen. Allgemein gilt, dass bei einem zu schadhafte Zustand des Originals in allen Einrichtungen eher vom Film digitalisiert wird. Wenn das Original nicht zur Verfügung steht, soll die Digitalisierung also „immer so original wie möglich“ (Ruff, SSB, I-2.1.) erfolgen.

In allen Einrichtungen finden Maßnahmen zur Qualitätssicherung der Digitalisate statt, wobei Werkzeuge wie ColorChecker, Auflösungsmaße oder Zentimetermaße zum Einsatz kommen. Die Kontrolle erfolgt dabei doppelt, das erste mal beim Erstellen der Images und danach nochmals durch eine eigene Qualitätssicherungsstelle (SBB) oder bei der Weiterverarbeitung in der Datenbank (SSB: Kontrolle durch die Fotothek auf Farb- oder Auflösungsfehler, falsche Farbprofile von Datenbank zurückgewiesen) oder der Metadatenerfassung (UB: bibliothekarisches Fachpersonal erstellt die Metadaten und kontrolliert Bilder und Paginierung). Bei den Bibliotheken wird zudem speziell auf Paginierung und Vollständigkeit der Bücher geachtet, was bei Museumsobjekten, wenn es sich nicht um Bücher handelt, weniger gegeben ist. Bei der SSB findet zudem eine laufende Kontrolle bei Bildanfragen statt, ob die Digitalisate noch aktuellen Qualitätsanforderungen entsprechen.

Hier zeigt sich, dass die Qualitätssicherung in die Digitalisierungsworkflows aller untersuchten Einrichtungen umfangreich eingebunden ist, die Maßnahmen sind dadurch dichter und routinierter als bei vielen anderen Forschungsdatenformen.

Eine nachträgliche Veränderung von publizierten Digitalisaten findet in der Regel in keiner der Einrichtungen statt, es wird Wert darauf gelegt, dass die Daten mit dem Ende des Digitalisierungsworkflows eine gute Qualität möglichst nah am Original erreicht haben. Die Daten sind somit persistent und für die Nutzung als Forschungsdaten geeignet. Eine Ausnahme, die in diesem Zusammenhang kritisch gesehen werden kann, sind Korrekturen von nachträglich entdeckten Fehlern wie fehlenden Seiten oder Bildfehlern, welche ersetzt und nicht dokumentiert/versioniert werden. Da dies jedoch keine inhaltlichen Veränderungen sind und Fehler bei der Nutzung als Forschungsdaten auffallen sollten, können diese Veränderungen hierbei vernachlässigt werden.

5.3.3.3. Dokumentation, Metadaten und Erschließungsstandards (I-3.).

Bei der Erschließung und den Metadatenstandards zeigen sich die Unterschiede zwischen Bibliothek und Museum besonders deutlich. Grundlegende Metadaten wie Titel, Autor/Künstler, Erscheinungsort und -jahr, Schlagworte sowie ein Identifier/Inventarnummer sind zwar überall vertreten, in der SBB und der UB werden die Metadaten jedoch über den regulären Katalog erfasst und dort bibliothekarisch erschlossen, weshalb Normdaten wie die GND oder klassische bibliothekarische Thesauri wie der Thesaurus für alte Drucke des GBV Verwendung finden. Hinzu kommen Strukturdaten, der Schwerpunkt liegt hier also auf der inhaltlichen Erschließung.

Beschrieben werden die Daten nach METS/MODS, in der SBB werden zudem die EXIF-Metadaten der Bilder ausgelesen.

Im Museum dagegen liegt der Schwerpunkt viel stärker auf der Objektbeschreibung, neben den grundlegenden Metadaten werden hier auch Material, Technik und Maße, Stempel und Markierungen, Daten zur Provenienz und Objektgeschichte, zur Restaurierung (Zustand, Lagerungshinweise, ...), Registrardaten und Hinweise zur Objektverwaltung (Standort, Leihstatus, Transportmöglichkeiten, ...) vermerkt. Auch objektspezifische Daten wie der Nominalwert einer Münze können erfasst werden. Erfasst werden die Daten in der nach dem Museumsstandard SPECTRUM zertifizierten Objektdatenbank Daphne, welche auch Normdaten nach der GND, GeoNames oder dem Art and Architectural Thesaurus (AAT) unterstützt und EXIF-Daten aus den Bildern ausliest.

Während die Metadaten bei den beiden untersuchten Bibliotheken frei in den jeweiligen digitalen Sammlungen zur Verfügung stehen, bietet die Sammlung Online der SSB nur die grundlegenden Metadaten an, alle weiteren Daten können nicht auf der Weboberfläche dargestellt werden. Auf Anfrage wird jedoch ohne weiteres Auskunft erteilt. Bei den beiden bibliothekarischen Angeboten lassen sich die Metadaten zudem als Dublin Core Datensätze herunterladen.⁵⁰ Dies entspricht den Empfehlungen, die für Forschungsdaten in den Geisteswissenschaften gegeben wurden (vgl. Puhl et al. 2015, S. 28).

Die Volltexterschließung findet regelmäßig an der UB, projektbezogen an der SSB und SBB statt, wobei die beiden Bibliotheken dabei ALTO-XML verwenden (die SBB zudem TEI und FineRead), die Texte in der SSB in die Datenstrukturen der Datenbank eingebunden sind und zusätzlich manchmal als Word-Dokumente vorliegen.

Der Kontext der Digitalisierung eines Objektes als wichtiges Element der wissenschaftlichen Nachnutzung wird in allen untersuchten Einrichtungen anders dargestellt. Während die SBB den Kontext durch Beschreibung des Projekttitels direkt im jeweiligen Objektdatensatz verfügbar macht, wird der Kontext in der UB durch eine Sammlungszuordnung gegeben. In der SSB finden Objektkennzeichnungen der Datenbank Anwendung, welche zwar das Browsen innerhalb der Objektes eines Digitalisierungsprojektes, nicht jedoch eine direkte Suche ermöglichen.

Auch der inhaltliche Kontext und die Provenienz werden unterschiedlich dargestellt. In der SSB gehört beides nicht zum Mindestdatensatz, ist also nicht im digitalen Angebot verfügbar. Eine Erfassung findet zudem nur projektbezogen bei einer wissenschaftlichen Bearbeitung statt, gerade bei Massendigitalisierungsprojekten fehlen tiefere Angaben daher oft.

In der SBB findet die Provenienzrecherche routinemäßig nur für Bestände vor 1850 statt, ansonsten wird der Provenienzvermerk zur Darstellung des (gegenwärtigen) Sammlungszusammenhanges verwendet. Die Angaben finden sich jedoch nicht im Metadatenatz, sondern nur im Katalog und dem speziellen Provenienz-Wiki.

An der UB sind Übersichtsseiten zu den Sammlungen bereits vorhanden, in Planung sind jedoch weitere Beschreibungen der Sammlungen und Digitalisierungsprojekte. Auch

⁵⁰ Siehe hierzu auch Abschnitt 5.3.1.2. Auswertung Webevaluation.

Hinweise zu Sammlung und Digitalisierungsprojekt sowie Zusatzinformationen an jedem Objektdatensatz sind aktuell noch in Planung.

Zu der Frage, ob umfangreiche Metadaten einen Ausgleich des Digitalisates zum Original bieten findet sich bei allen Gedächtniseinrichtungen eine klare Meinung. Ja, viele Metadaten, gerade auch in Spezialfällen wie Einbänden, Bundpapieren/Vorsatzpapieren oder Wasserzeichen, können durchaus einen gewissen Ausgleich schaffen, werden es aber (mit aktuellen technischen Mitteln) niemals schaffen, die volle Repräsentanz des Originals inklusive Haptik oder Größenwahrnehmung wiederzugeben. Das Ersetzen des Originals ist jedoch auch nicht das Ziel, sondern vielmehr die Sichtbarmachung der Bestände. Zudem bieten umfangreich erschlossene Digitalisate massive Zusatznutzen für die Nutzer, von denen sich die meisten nur für den Inhalt des Buches interessieren. Gut ausgestattete Digitalisate ermöglichen hier die umfangreichen Analysemethoden der Digital Humanities. Durch neue Aufnahmeverfahren können zudem auch bisher nicht zugängliche Informationen sichtbar gemacht werden, wie z.B. die Digitalisierung einer nicht mehr zu öffnenden Schriftrolle, indem die eisenhaltige Tinte mittels computertomographischer Verfahren gemessen wird. Überhaupt werden durch die Digitalisierung zuvor lange verschlossene Bestände sichtbar gemacht, was hier durchaus als ein Demokratisierungsprozess zu betrachten ist.

Diese Zusatznutzen rechtfertigen es jedoch nicht, das Objekt an sich auszusortieren, da bestimmte medientypische Informationen nicht übertragen werden können. Hinzu kommt, dass nicht nur solche medientypischen Informationen, sondern auch die Unterschiede von mehrfach hergestellten Objekten wie z.B. alten Drucken wichtige Forschungsaspekte sind, die allein an Hand von Digitalisaten nicht oder nur unzureichend möglich sind.

5.3.3.4. Veröffentlichung (I-4.).

Bei den verwendeten Lizenzen zeigen sich klare Nutzungsunterschiede. Die freiesten Lizenzen werden bei der SBB verwendet, wo der Standardfall die Public Domain ist. Dazu kommen die drei Sonderfälle CC-Lizenzen, Copyright und eigens verhandelte Lizenzen. Die Public Domain Lizenz, welche auch für die Katalogeinträge und Metadaten gilt, hat den Vorteil, dass die Nutzer sie für ihre Zwecke frei verwenden können, während z.B. eine CC-NC Lizenz auch die Nutzung der Digitalisate bei Wikipedia verhindern würde. Allerdings entfällt somit auch die Namensnennung, welche eine gute Art der Repräsentation des Hauses darstellen würde. Herr Siegmann nennt hierbei als Beispiel die französische Gallica, die an ihre freien Bilddigitalisate einen schmalen weißen Balken mit der Namensnennung angefügt hat. Dieser kann ohne Probleme entfernt werden, oftmals wird dies jedoch vergessen und somit die Herkunft des Bildes sofort sichtbar.

In der UB wird dagegen die CC-BY-SA-NC Lizenz als Standard verwendet, nur einige Digitalisate stehen durch Verträge mit den Rechteinhabern noch unter anderen Lizenzen. Durch den hier verwendeten Standard ist die Namensnennung garantiert und die kommerzielle Nutzung kann zudem gesteuert werden. Damit wird eine kostenfreie kommerzielle Nutzung nach Absprache nicht verhindert, es ist jedoch gewünscht, dass vorher Kontakt aufgenommen wird.

Die SSB verwendet dagegen, noch aus traditionellen Entwicklungen, gerade im Museumsbereich, heraus, eine Copyright-Regelung. CC-0 oder CC-BY Lizenzen kommen als Ausnahmen auch vor, nämlich für alle Objekte, die im Museum-Digital oder der DDB verwendet wurden. Durch die Copyright-Regelung ist eine Kontrolle und Vermarktung der Bildbestände möglich, wobei gerade im wissenschaftlichen Bereich auf eine Kostenerhebung verzichtet wird. Zudem wird so die Namensnennung der Reproduktionsfotos garantiert. Aktuell wird allerdings der Umstieg auf eine offene CC-Lizenz angestrebt, da dies Klarheit bei der Nachnutzung schafft und zudem eine sichtbare Umsetzung der Open Access Strategie des Hauses darstellt.

Insgesamt zeigt sich also, dass alle der untersuchten Einrichtungen offene Lizenzen verwenden oder diese anstreben, die Nachnutzung der Digitalisate also vollkommen gewünscht ist.

Im Rahmen der Lizenzvergaben gibt es daher auch keine bzw. keine weiteren Zugangsbeschränkungen zu den publizierten Digitalisaten. Durchaus ist jedoch das Vorhandensein von unpublizierten Digitalisaten möglich, welche sich noch im Workflow befinden. In der SSB beträgt die Rate hier etwa 1:5, also 300.000 Datensätze zu 60.000 Objekten in der Sammlung Online. Grund dafür sind oft noch fehlende Bilder oder die noch nicht abgeschlossene Qualitätskontrolle, da Metadaten und Bilder einen gewissen Stand haben sollen, bevor sie veröffentlicht werden.

Durch die unterschiedlichen Lizenzen sind die Daten in der SBB und der UB, zumindest für wissenschaftliche Nutzer, frei zugänglich, in der SSB dagegen nur sichtbar. „Die Digitalisate sind nicht in hoher Auflösung herunterzuladen, die Metadaten sind nicht über eine API oder ein Repositorium zugänglich. Wir sind sichtbar, die Daten sind aber nicht frei zugänglich“ (Ruff, SSB, I-4.1.).

Dementsprechend sind in der Sammlung Online des Stadtmuseums die TIFF-Digitalisate auch nicht herunterzuladen. Viele Datensätze haben zudem kein TIFF-Bild. Dies liegt zum einen daran, dass nicht genug Ressourcen vorhanden sind, um von allen Objekten sofort hochwertige Bilder anzubieten und daher auf JPEG-Arbeitsfotos von Mitarbeitern aus den Sammlungen zurückgegriffen wird. Zum anderen machen TIFF-Abbildungen nicht bei allen Objekten Sinn, so dass Speicherplatz in der Sammlung Online gespart werden kann, auch wenn dadurch ein Auswahlprozess stattfindet, bei welchen Objekten ein hochauflösendes Bild nützlich ist.

Auch die beiden Bibliotheken bieten in ihren digitalen Angeboten nicht die höchstmögliche Qualität, sondern JPEG-komprimierte TIFF-Dateien an, welche zwar für die meisten Gebrauchsanwendungen bis hin zu Präsentationen oder Webpublikationen sehr gut zu nutzen sind, auf Grund der geringeren Größe zudem Vorteile beim Speicherplatz und der Übertragungsgeschwindigkeit bieten.

Trotz der unterschiedlichen Lizenzen bei den publizierten Digitalisaten gelten doch in allen drei Einrichtungen gleiche Regelungen für den Erhalt von hochaufgelösten Master-Dateien. Durch Anfrage oder DoD-Auftrag ist es möglich, die Master-Dateien zu erhalten, jedoch muss in allen Einrichtungen dabei eine Bearbeitungsgebühr gezahlt werden.

Allerdings kann so, wie z.B. in UB und SSB gewünscht, die korrekte Zitierweise des Digitalisates vorgegeben werden, es findet also eine Kontrolle auf Richtigkeit statt.

Wie auch in der Literatur beschrieben (vgl. Münster et al. 2018, S. 379) werden in den hier untersuchten Einrichtungen gut aufgelöste Daten angeboten, die für einen Großteil der Forschungsanwendungen, gerade im meist inhaltlich arbeitenden historischen Bereich, vollkommen ausreichen.

Ziel der jeweiligen digitalen Angebote ist es, wie aus den Lizenzen hervorgeht, die Digitalisate nachnutzbar, oder zumindest sichtbar, und auffindbar zu machen. Damit bilden sie als Repositorien von Digitalisaten eine Art von Forschungsdatenrepositorium, welches die jeweilige Einrichtung für ihre Nutzer bereitstellt. Die sonst üblichen Anforderungen an solche Repositorien wie das DINI-Zertifikat oder das Data Seal of Approval spielen hierbei jedoch keine Rolle, keines der Angebote hat ein solches Zertifikat bisher erhalten oder prüfen lassen. An der SBB gibt es allerdings Bestrebungen, für die Born Digitals- und FID-Angebote solche Zertifikate zu erreichen, für ihre Digitalisierten Sammlungen wäre es in diesem Zusammenhang auch wünschenswert.

Zur Qualitätssicherung der eigenen Angebote scheint man weniger auf externe Prüfungen als auf eigene Kontrolle zu setzen, was jedoch auch an der Reputation der untersuchten Einrichtungen liegen kann, die ein zusätzliches Qualitätszertifikat überflüssig zu machen scheint.

5.3.3.5. Interoperabilität und Schnittstellen, Portale (I-5.).

Klare Unterschiede zwischen den Einrichtungen zeigen sich auch bei den Schnittstellen und teilweise bei den bespielten Portalen.

Die Sammlung Online der SSB ist eine klassische digitale Präsentationsplattform für Museen, es werden den Nutzern daher keine Schnittstellen oder Harvestingmöglichkeiten angeboten. Allerdings kann die dahinter liegende Daphne Datenbank in das museumsspezifische LIDO-Format exportieren, welches Nutzern auf Anfrage bereit gestellt werden kann. Anders ist dies in den Bibliotheken, die traditionell schon früh auf Datenaustausch und damit Schnittstellen gesetzt haben, da dies sich bei Büchern eher anbietet als bei unikalenen Museumsobjekten. Dementsprechend bieten die Angebote der SBB und der UB sowohl verschiedene Exportformate als auch verschiedene Schnittstellen zum Harvesting von Daten an. Während die UB eine OAI-PMH-Schnittstelle anbietet, stellt die SBB ihre Daten neben dieser auch über eine IIIF-Schnittstelle oder eine SRU-Schnittstelle des GBV für Katalogdaten bereit.⁵¹

Die Deutsche Digitale Bibliothek (DDB) wird als zentrales Portal der deutschen Gedächtnisinstitutionen und nationaler Aggregator der Europeana von allen drei untersuchten Einrichtungen bespielt. Die DDB bietet daher eine gute Chance, Objekte über Institutionsgrenzen hinaus sichtbar zu machen. Ein Hochladen der Digitalisate in die DDB ist aber oftmals auch Anforderung von Projektförderern wie digiS (dem Forschungs- und Kompetenzzentrum Digitalisierung des Landes Berlin) oder der DFG (vgl. DFG 2016,

⁵¹ Sieh auch Abschnitt 5.3.1.2. Auswertung Webevaluation. Dass die OAI-PMH- und IIIF-Schnittstelle der SBB hier nicht aufgeführt ist, ist dem Umstand geschuldet, dass diese noch nicht aktiv auf der Website beworben werden, aber dennoch aktiv sind.

S. 31). Dabei berichten alle drei Institutionen von aktuellen technischen Problemen, die Daten aus der DDB weiter in die Europeana zu spielen, so dass hier bisher kaum Daten vorhanden sind, obwohl die Europeana unterstützt wird. Die weiteren Portale unterscheiden sich dann je nach Einrichtungstyp. Während die SSB ihre Daten (zumindest projektweise) auf der Plattform Museum-Digital einspielt, welche auch gemeinsame Projekte mit anderen Einrichtungen erlaubt und gerade für den Berlin Fokus besonders wichtig ist, spielen UB und SBB ihre Daten in das Zentrale Verzeichnis Digitalisierter Drucke (ZVDD). Im Stadtmuseum kommen zudem einige „Altlasten“ in Spezialdatenbanken, die von (ehemaligen Sammlungskuratoren) eingespielt, aber nicht weiter gepflegt oder zentral erfasst worden sind sowie Datensätze der Bibliothek und Dokumentensammlung in Kalliope, KOBV und ZDB.

Eine Verlinkung im Sinne von Linked Open Data, welche zusätzliche Kontextinformationen oder weitere Zusammenhänge zwischen Objekten ermöglichen würde, findet in keiner der Institutionen statt.

Der Großteil der Verlinkungen entsteht über die Verwendung der selben normierten Begriffe in den Metadaten, hinzu kommen Gruppierungen wie digitale Kollektionen oder Projektzusammengehörigkeiten. In der SSB bietet die Datenbank zumindest Objektzusammenbeziehungen wie „ist gleichrangig“ oder „ist Teil von“ an, was jedoch weder stark genutzt noch in der digitalen Darstellung sichtbar wird. Die beiden Bibliotheken verlinken zwischen ihren Objekten in den digitalen Angeboten und dem OPAC.

Eine richtige Verlinkung zwischen den Objekten im Sinne von Objektnetzwerken und Linked Data ist also nicht gegeben.

Ein direktes Angebot auf Wikipedia gibt es nur von Seiten der SSB, da man hier im Kontakt mit Wikimedia steht. Die Inhalte sind dabei aber vor allem Ergebnisse von Hackathons oder verfasste Beiträge, Digitalisate werden kaum in die Wikipedia gespielt. Von den anderen beiden Häusern gibt es zwar auch Digitalisate in der Wikipedia, diese wurden jedoch von Nutzern und nicht durch die Institutionen hochgeladen.

Alle drei Institutionen nutzen zudem Social Media Kanäle, wobei die SSB hier vor allem Öffentlichkeitsarbeit betreibt, Digitalisate werden für redaktionelle Zwecke, nicht jedoch als selbstständige Inhalte genutzt. In der SBB sind die Digitalisate anlassbezogen auf den Social Media Kanälen vertreten, wie z.B. bei der Vorstellung prominenter Digitalisate wie den Humboldt-Tagebüchern. Einzig in der UB hat die Historische Sammlung als für die Digitalisierung zuständige Abteilung einen eigenen Social Media Account bei Flickr, auf welchem regelmäßig neue Digitalisate publiziert werden, welche vielleicht wissenschaftlich nicht allzu interessant, dafür aber attraktiv sind und über diesen Weg andere Gruppen als die typischen Nutzer ansprechen sollen.

Trotz der Möglichkeiten wird die Publikation von Digitalisaten in der Wikipedia im Sinne einer Wissenspopularisierung nicht genutzt, statt dessen sind die Einrichtungen mit ihren Digitalisaten jedoch mehr oder weniger stark auf Social Media Plattformen vertreten. Das eine starke Öffentlichkeitsarbeit direkt aus den Digitalisierungsabteilungen, die durch ihre Querschnittsfunktion einen guten Überblick über alle Sammlungen des Hauses haben,

sehr sinnvoll und effektiv sein kann, wird dabei in der SBB durchaus gesehen, scheitert hier jedoch an der Größe des Hauses, da es zu aufwändig wäre, zu jedem digitalisierten Objekt die fachlichen Experten nach möglichen Besonderheiten zu fragen, so dass solch eine Umsetzung allein in der UB realisiert wurde.

5.3.3.6. Speicherung, Pflege und Archivierung (I-6.).

Obwohl die Langzeitarchivierung ein wichtiger Teil des (Forschungs-)Datenmanagements ist, besitzt keine der drei untersuchten Institutionen bisher eine Strategie zur Langzeitarchivierung, sondern es werden nur Backup-Strategien zur Speicherung verfolgt, welche in der Regel auf regelmäßigen, redundanten und örtlich getrennten Sicherungen beruhen. Dass das Thema dennoch wichtig ist, zeigen jedoch die Entwicklungen. Während die SSB ihre Langzeitarchivierung an das Zuse Institut Berlin als Anbieter der Langzeitarchivierung für Landeseinrichtungen Berlins ausgelagert hat, steht auch die UB in Kontakt mit diesem und hat zudem den Status eines Pilotpartners für das EWIG-Projekt zur Langzeitarchivierung. An der SBB wird dagegen gerade durch entsprechende Mitarbeiter eine eigene Langzeitarchivierungsstrategie entwickelt.

Spezielle Formate werden dabei nicht verwendet, alle Einrichtungen speichern ihre Daten in den Formaten, mit denen auch gearbeitet wird. Die SSB verwendet vereinzelt zudem RAW-Formate für die Speicherung von Bildern, an der SBB wird auf struktureller Ebene darauf geachtet, dass es keine verteilten Systeme gibt, jedes Digitalisat also mindestens ein Rumpfkatalogisat im zentralen Katalog hat.

Dass eine gute Langzeitarchivierung oder Speicherung sich auf jeden Fall lohnt, auch wenn die Digitalisate, anders als viele andere Forschungsdaten, neu erzeugt werden können, wird von allen Interviewten beschrieben. Die Kernargumente hierbei sind überall organisatorische Effektivität und Bestandserhaltung. Da die Kosten für die Erhaltung von Digitalisaten durch günstigere Speichermöglichkeiten kontinuierlich sinken, der Aufwand für den gesamten Digitalisierungsworkflow aber gleich bleibt, ist die Speicherung langfristig günstiger. Zudem stellt die Digitalisierung eine intensive Belastung der Bestände dar, welche möglichst nicht wiederholt werden soll.

Hinzu kommen weitere Argumente, wie die Zustandsdokumentation, die durch jedes Digitalisat automatisch erfolgt, die Mengen an überhaupt noch nicht digitalisierten Beständen, denen der Vorrang zu geben ist oder die Verpflichtung auf Persistenz, wenn Identifier wie PURL oder DOI verwendet werden sollen.

Allerdings wird auch überall beschrieben, dass sich Standards in Zukunft ändern können. So hätte man bei der Sicherungsverfilmung vor einigen Jahrzehnten auch nicht gedacht, die Bestände später ein weiteres mal aufnehmen zu wollen. Durch die Verwendung von Standards hofft man jedoch, dass mögliche Änderungen viele Institutionen betreffen und dann durch Ansätze wie Massenmigrationen möglichst ohne große Probleme durchgeführt werden können, eine neue Digitalisierung jedoch nicht nötig ist.

Wenn Digitalisate doch neu aufgenommen werden, so wird zumindest in den beiden Bibliotheken der Anspruch an Forschungsdaten erfüllt, diese nachvollziehbar und referenzierbar zu halten. In der UB ist dies bisher praktisch noch nicht vorgekommen, da

diese erst sehr spät mit der Digitalisierung begonnen hat. An der SBB dagegen beginnt gerade eine Ära, dass man durch technische Weiterentwicklungen und mehr Erfahrung im Umgang mit speziellen, z.B. reflektierenden Objekten, auch eine bessere Qualität bereits bestehender Digitalisate erreicht werden könnte. Eine eindeutige Referenzierung durch Versionierung und persistente Adressierung ist daher wichtig, in der Praxis werden bei solchen Fällen die neuen Digitalisate in den Digitalisierten Sammlungen angezeigt, bestehende PURLs/DOLs verweisen aber weiterhin auf ihre ursprünglichen Versionen. Die SSB erfüllt diesen Anspruch indirekt, da hier die Erhaltung aller Digitalisate als museologische Zustandsfotos im Vordergrund steht, dementsprechend z.B. ältere JPEGs nicht für ein neues TIFF gelöscht werden.

Diesen Ansätzen folgend ist in den untersuchten Einrichtungen auch nicht geplant, veraltete Formate zu löschen. In der Praxis stellte sich diese Frage jedoch noch nicht, da alle Institutionen nur Formate verwenden, die bisher noch lesbar sind. An der SBB gibt es zwar teilweise Diskussionen, auf Grund von Speicherplatzeffektivität von TIFF auf JPEG2000 umzustellen, in diesem Fall würden die Bestände jedoch konvertiert und die TIFF-Digitalisate als Sicherung aufbewahrt werden.

Probleme anderer Forschungsdaten wie z.B. proprietäre Formate oder fehlende deskriptive Metadaten (vgl. Pempe 2012, S. 151) stellen sich hier also auf Grund von Standardformaten und klaren Workflows bei der Erstellung nicht.

5.3.3.7. Organisatorisches (I-7.).

Förderungen der Digitalisierung erhalten die untersuchten Einrichtungen von unterschiedlichen Stellen mit Ausnahme der DFG, die alle Einrichtungen projektweise fördert oder zumindest schon gefördert hat. Die SBB als Bundeseinrichtung erhält neben der DFG Förderungen durch das Bundesministerium für Bildung und Forschung (BMBF) und den Beauftragten der Bundesregierung für Kultur und Medien (BKM), zudem durch die Europäische Union. Die SSB als Einrichtung des Landes Berlin erhält dagegen über digiS Projektmittel des Landes Berlin, hinzu kommen europäische EFRE-Mittel. Die UB ist als Teil der Humboldt-Universität auch eine Landeseinrichtung, da Universitätsbibliotheken vom Land Berlin jedoch nur in zweiter Linie gefördert werden, waren Projektanträge hier bisher noch nicht erfolgreich.

Eine Weiterentwicklung der jeweiligen digitalen Präsentationssysteme findet in allen Einrichtungen statt, neue Funktionen werden regelmäßig hinzugefügt. Während SSB und UB dies über Dienstleister beauftragen, werden Updates in der SBB durch die eigene IT-Abteilung realisiert. Die Weiterentwicklung der Digitalisierten Sammlungen geschieht über die AG Präsentation, welche neue Entwicklungen plant und umsetzt. Zudem bietet die SBB einen Zugriff auf eine offene Beta-Version des Präsentationssystems, so dass neue Funktionen schon vor dem eigentlichen Release von Nutzern ausprobiert werden können.⁵²

⁵² Staatsbibliothek zu Berlin. Digitalisierte Sammlungen Beta-Version, <https://digital-beta.staatsbibliothek-berlin.de/> (abgerufen am 30.07.2019, 15:38).

Ein Datenmanagementplan als wichtiges Verwaltungsinstrument für Forschungsdaten liegt konkret in keiner Einrichtung vor, die Aspekte des Datenmanagements werden aber durch die hier beschriebenen Standards, Workflows und Speicherverfahren abgedeckt und sind somit fester Bestandteil der Aufgaben in den Häusern. Hinzu kommt, dass alle untersuchten Gedächtnisinstitutionen feste öffentliche Einrichtungen sind, so dass die Gefahr eines Verwaisens von Daten, wie es oft nach dem Ende von Forschungsprojekten der Fall ist, hier sehr gering ist.

Bei den Zielgruppen zeigen sich die unterschiedlichen Profile der Einrichtungen besonders deutlich. Die SSB hat ein sehr breites Zielpublikum, was sich aus der interessierten Öffentlichkeit, Kollegen anderer Museen sowie, besonders in speziellen Sammlungen, Forschung und Lehre zusammensetzt. Hinzu kommen auch kommerzielle Nutzer bei der Bildverwertung. Die SBB legt sich als Universalbibliothek zwar auch nicht auf nur eine Nutzergruppe fest, insgesamt würde man aber wahrscheinlich der Wissenschaft, inklusive Laienforschern, den Vorzug geben. Eine Ausnahme bilden die FIDs, wo die jeweiligen Fachcommunities klar den Schwerpunkt bilden. Die UB dagegen hat als Hauptzielgruppe ganz klar die Wissenschaft, besonders an Universitäten, wozu auch das Studium als Ausbildung für die Wissenschaft zählt. Bestandskonzept und Bestandsmanagement und auch die Digitalisierung richten sich daher an der wissenschaftlichen Nutzung, mit besonderem Fokus auf die Wissenschaft an der Humboldt-Universität, aus. Die interessierte Öffentlichkeit wird natürlich nicht von der Nutzung ausgeschlossen, stellt aber nur am Rande eine Zielgruppe dar.

Offener sind dagegen die Sammlungspräsentationen ausgerichtet. Eine konkrete Erhebung der Nutzergruppen gibt es allerdings nicht, es werden nur mittels Tools wie Google Analytics oder Matamo (ehemals Piwik) Nutzerstatistiken erstellt.

Eine allgemeine Einbeziehung der Nutzer in die Digitalisierungsplanung findet nicht statt, es sei denn, es bestehen direkte Kontakte zu diesen Nutzern. So berücksichtigt die SBB z.B. Kontakte der Fachabteilungen und FIDs zu ihren jeweiligen Communities und in der UB werden bei Beständen mit spezifischen Nutzergruppen wie z.B. den Gelehrtenbibliotheken die Nutzer zu ihren Prioritäten im jeweiligen Bestand befragt.

Eine Verknüpfung von Digitalisaten und daraus hervorgegangenen Forschungsdaten findet bei allen Interviewten Zustimmung, ein Hosting solch eines Angebotes auf Seiten der Gedächtniseinrichtung wird jedoch eher problematisch gesehen.

In der SSB wird darauf verwiesen, dass man kein weiteres Portal für die Fachcommunity aufbauen, sondern die Daten in Infrastrukturen zusammenführen sollte, die bereits stark von der Forschung genutzt werden. Das eigene System sollte dann mit Links auf die Objekte dort ausgestattet werden.

Bei der UB wird hier zudem der Vorteil gesehen, dass die Daten so nachhaltig gesichert und verknüpft werden könnten, was bei den Forschungstools oft nicht gegeben ist. Für eine Umsetzung im Haus fehlt hier aber sowohl die Infrastruktur wie der Auftrag der Universität, eine solche einzurichten.

Die SBB sieht hierbei einen Vergleich zu den analogen Belegexemplaren, wobei die praktische Umsetzung an den hier verwendeten offenen Lizenzen scheitert. Dadurch,

dass die Digitalisate frei zugänglich und nachnutzbar sind und die Daten zudem über Schnittstellen angeboten werden, wird nicht verfolgt, wer die Daten nutzt. Man kann hier daher nur mit der Bitte, z.B. durch einen Hinweistext, an die Nutzer herantreten, aus den Digitalisaten entstandene Forschungsdaten einzureichen, damit diese verknüpft werden können.

Trotz der unterschiedlichen Einrichtungen haben die Interviewten relativ ähnliche Vorstellungen davon, was sich in Gedächtniseinrichtungen verbessern sollte bzw. könnte. Von allen Befragten wird dabei eine bessere Zusammenarbeit mit den Nutzern genannt. Dies bezieht sich zum einen auf die zu digitalisierenden Bestände, vor allem aber auch auf die von den Nutzern erwarteten und benötigten Funktionen. Gerade für die digitalen Sammlungen sollte hier weniger vom klassischen Bibliotheksnutzer ausgegangen werden, der nur lesend auf die Bestände zugreifen möchte, sondern es sollten auch die Möglichkeiten der Digital Humanities berücksichtigt werden. Im Zuge dessen könnte auch an eine Anpassung der Präsentationsschicht, gerade für jüngere Nutzer, die hier anderes gewohnt sind, stattfinden. Als wichtig wird mehrmals auch die freie Verfügbarmachung im Sinne von Open Access für gemeinfreie Werke durch die Gedächtniseinrichtungen beschrieben, da dies Wissenschaft und Allgemeinheit zugute kommen und die Nutzung vereinfachen würde.

Als weitere Verbesserungsaspekte werden einheitliche Qualitätsstandards und Workflows für möglichst alle Gedächtniseinrichtungen und die Zusammenarbeit mit Portalen genannt, statt für jedes Projekt ein eigenes Portal aufzubauen (SSB). In diese Richtung weist auch der Vorschlag, Angebote wie die DDB oder Europeana noch stärker auszubauen, um die hier liegenden großen Potentiale voll auszuschöpfen (SBB).

In der SSB wird auch der Aufbau von eigenen Ressourcen zur Digitalisierung als wichtige Verbesserung gesehen. Da die Digitalisierung inzwischen Kernaufgabe von Gedächtniseinrichtungen ist, darf sie nicht vom Erfolg einzelner Förderprojekte abhängig, sondern muss als Gesamtauftrag der Institution in allen Bereichen bis hin zur Geschäftsleitung verankert sein. Dazu werden die entsprechenden Personalstellen und Techniken benötigt.

Insgesamt wird, auch in anderen Teilen des Interviews, klar, dass viele Weiterentwicklungen und neue Angebote möglich sind, dies jedoch auch immer eine Frage der personellen und finanziellen Mittel ist.

Dieser Aspekt wird daher auch in der letzten Frage des Interviews thematisiert. So beschreiben alle Interviewten, dass ihre Institutionen bei unbegrenzten Mitteln viel schneller, viel tiefer und viel mehr digitalisieren und erschließen könnten und der gesamte Bestand der jeweiligen Einrichtung innerhalb weniger Jahre online verfügbar und nutzbar wäre. Außerdem wird von allen beschrieben, dass man die digitalen Sammlungen so viel besser mit internen wie externen Sammlungen verknüpfen könnte. Man würde entweder die eigene Sammlung auf Portalen noch sichtbarer machen und so in der Öffentlichkeit viel präsenter sein sowie neue Nutzergruppen erschließen oder man würde selber zu einer zentralen Anlaufstelle für bestimmte Themen werden, indem man einen Einstiegspunkt in diese Themen, Informationen und Sachverhalte bietet und Bezüge unter

Objekten herstellt anstatt nur Materialien zu liefern. Egal ob auf Portalen oder eigenen Angeboten, die jeweilige Einrichtung könnte so (noch mehr) zu einer verlässlichen Quelle und Anlaufstelle für Wissen und Daten im Internet werden, besonders auch bei Zielgruppen, die die Angebote bisher nicht nutzen.

Auch könnten die eigenen Konzepte weiterentwickelt werden. In der UB könnte man z.B. universitätsweit sammlungsübergreifend arbeiten und passende Objekte in die Digitalisierung mit einbinden, um daraus digitale Ausstellungen zu kuratieren. Auch die Angebote wie der Viewer ließen sich verbessern und mit neuen Funktionen ausstatten.

Mit dem Abschluss der Digitalisierungsarbeiten würden in der SSB auch wieder Kapazitäten für andere Museumsaufgaben frei werden, so dass man sich wieder mehr auf die Forschung im Haus konzentrieren könnte. Die Digitalisate ließen sich auch viel besser in neue Formen der Museumsarbeit einbinden und man hätte die Mittel, Croudsourcing-Angebote umzusetzen, um die Community in die Erschließung von Objekten mit einzubinden.

5.4. Welche Anforderungen haben Nutzer an Digitalisate als Forschungsdaten?

5.4.1. Beschreibung der Fragen und des Aufbaus (Methodik) - Nutzer.

Wie beschrieben orientieren sich die Fragen am Arbeitsablauf für die Nutzung von Digitalisaten, wobei erst auf das zu nutzende System, dann die Anforderungen an die Digitalisate und ihre Verfügbarkeit und Nachnutzung eingegangen wird. Abschließend wird nach den Vorstellungen und Wünschen der Nutzern an die Digitalisate sowie die Einrichtungen gefragt.

Das Interview beginnt mit einer Faktenfrage zum System von digitalen Sammlungen, wobei die Nutzer angeben sollen, ob ihnen als Sucheinstieg die Suche oder Browsing und Facettierung wichtiger sind (Nf-1.1.). Dadurch soll geklärt werden, ob die Interviewpartner eher mit konkreten Suchen oder eher durch kontinuierliche Facettierung arbeiten.

Auch die weiteren Fragen zum favorisierten System dienen der Charakterisierung der Nutzer, indem sie beschreiben, ob sie eher Portale wie die DDB oder die Europeana oder direkt die digitalen Angebote von Gedächtniseinrichtungen nutzen (N-1.1.). Diese Frage soll zudem einen Hinweis für Gedächtniseinrichtungen geben, welche Art des Angebots zu bevorzugen ist und wird daher mit der entsprechenden Frage auf der Institutionenseite abgeglichen. Auch die darauf folgende Frage, nämlich welche Funktionen am meisten durch die Nutzer verwendet werden und am wichtigsten für sie sind (N-1.2.), soll den Institutionen einen Hinweis geben und wird daher mit dem aktuellen Angebot der untersuchten Gedächtnisinstitutionen, welches in der Webevaluation untersucht worden ist, verglichen.

Die Faktenfragen des darauf folgenden Abschnitts zu den Anforderungen der Nutzer an die Digitalisate konzentrieren sich auf die Metadaten, welche auf jeden Fall und welche idealerweise gebraucht werden (Nf-2.1.) und die Grundanforderungen der Nutzer an Bilddigitalisate wie z.B. Farbigkeit, Vergrößerungsfunktion oder Volltexterschließung (Nf-2.2.). Auch diese beiden Fragen decken sich mit Fragen an die Institutionen und sollen so später abgeglichen werden.

Zudem sollen Nutzer erklären, mit welchen digitalen Daten sie hauptsächlich arbeiten und daher von Gedächtniseinrichtungen benötigen (N-2.1.) und ob sie bestimmte Metadateformate und -schemata oder Normdaten für die Erschließung favorisieren (N-2.2.), um diese mit den von den Institutionen angebotenen Formaten abgleichen zu können. Mit den Ergebnissen aus den Institutioneninterviews sollen auch die folgenden Fragen verglichen werden: Zum einen, welche Informationen die Nutzer von einer Dokumentation eines Digitalisates erwarten (N-2.3.), zum anderen, welche Verarbeitungsstufe für die Nutzer bei Text- bzw. Bilddigitalisate am sinnvollsten ist (N-2.4.). Die Fragen sollen klären, wie viele Informationen sie brauchen und bis zu welchem Grad diese aufbereitet sein sollten. Interessant ist in diesem Zusammenhang auch, wie wichtig den Nutzern eine Verlinkung der Objekte untereinander ist (N-2.5.), da dies leichter zum Auffinden weiterer passender Objekte und Kontextinformationen führen könnte. Abgeschlossen wird dieser Bereich durch zwei Meinungsfragen, die in gleicher Weise auch den Interviewpartnern aus den Institutionen gestellt werden. Zum einen sollen die Nutzer ihre Erwartung beschreiben, wie Gedächtniseinrichtungen mit Dubletten umgehen sollten (N-2.6.), um zu klären, ob ihnen eher der Inhalt oder das Objekt an sich wichtig ist, da dies relevant für Forschungsfragen sein kann. Zum anderen sollen sie ihre Meinung zu der Idee nennen, dass viele Metadaten bei der Beschreibung von Digitalisaten den Verlust von Informationen des analogen Originals ausgleichen oder das Digitalisat gar das Original für die Forschung ersetzen kann (N-2.7.).

Die Fragen des Abschnitts zur Verfügbarkeit und Nachnutzung der Digitalisate bauen aufeinander auf, so dass hier keine Faktenfragen gestellt werden. Zum Einstieg in diesen Abschnitt sollen die Interviewpartner daher kurz beschreiben, ob die digitale Verfügbarkeit einer Quelle ihre Quellenauswahl beeinflusst (N-3.1.). Dadurch soll geklärt werden, ob sie die Digitalisate als Quellen oder nur zur Kontrolle analoger Quellen nutzen. Dies ist auch insofern interessant, da die bisherige historische Forschung durch den Zugang zu den (analogen) Quellen eingeschränkt war, ein Hindernis, welches durch digitale Verfügbarkeiten behoben werden könnte, wenn denn digitale Quellen von den Nutzern angenommen und akzeptiert werden. Daran schließt auch die Frage an, ob die Nutzer Digitalisaten aus Gedächtniseinrichtungen überhaupt als wissenschaftlichen Quellen vertrauen und ob dies an bestimmte Bedingungen geknüpft ist (N-3.2.), da Informationen aus dem Internet, wie bereits beschrieben, lange Zeit wegen fehlender Persistenz und Referenzierbarkeit nicht als geeignete Quellen angesehen wurden (vgl. Schubert 2018, S. 229). Weiterführend wird dann gefragt, wie das persönliche Forschungsdatenmanagement der Interviewten aussieht und ob sie sich dabei auf Angebote von virtuellen Forschungsumgebungen oder Cloud-Diensten stützen oder lokal arbeiten. Zudem soll beschrieben werden, welche Daten für die Arbeitsweise benötigt werden und ob es bestimmte Anforderungen von der Fachcommunity oder Forschungsförderern gibt (N-3.3.). Aus den Arbeitsweisen der interviewten Nutzer könnten Rückschlüsse für die Gedächtniseinrichtungen gezogen werden. Auf das persönliche Forschungsdatenmanagement baut die Folgefrage auf, die die Bedeutung der Downloadfunktion für die Nutzer beleuchten soll (N-3.4.). Diese stellt eine Besonderheit unter den Funktionen der digitalen Sammlungen dar, da Daten so auch lokal oder in

anderen Tools genutzt werden können. Eine Form der Weiternutzung der von den Gedächtniseinrichtungen zur Verfügung gestellten Daten kann das Harvesting sein, welches im Zuge der Digital Humanities zunehmend auch für quantitative Untersuchungen in der Geschichtswissenschaft interessant werden kann. Daher soll abgeglichen werden, ob die Befragten Harvesting zum Abrufen großer Datenmengen nutzen (N-3.5.) und Gedächtnisinstitutionen daher entsprechende Schnittstellen zur Verfügung stellen sollten. Abschließend sollen die Interviewpartner in einer Meinungsfrage kurz begründen, wie sie zum Thema Open Access für Digitalisate (sowohl bezüglich der Zugänglichkeit wie der Nachnutzung) stehen (N-3.6.), da Open Access wie beschrieben Zugänglichkeit sowie Nachhaltigkeit sichern und somit ein wichtiger Aspekt des Forschungsdatenmanagements sein kann.

Abschließend folgt eine Reihe an Meinungsfragen zu den Vorstellungen und Wünschen der Nutzer an die Digitalisate und Gedächtniseinrichtungen. Um hier die Wünsche und Möglichkeiten von Nutzern und Institutionen abgleichen zu können, wird die Frage, ob die Interviewten als Nutzer gern in Entscheidungsprozessen von Gedächtniseinrichtungen zur Digitalisierung einbezogen werden möchten (N-4.1.) gespiegelt auch auf Seite der Institutionen gestellt. Als nächstes sollen die Nutzer dann beschreiben, was für sie die Forschungsdaten der Geschichtswissenschaft sind (N-4.2.). Dadurch soll untersucht werden, ob sich diese Definition mit der genannten Forschungsdatendefinition deckt oder die Nutzer die Digitalisate vielleicht gar nicht als Forschungsdaten wahrnehmen. Die Frage ist dabei bewusst erst am Ende des Interviews gestellt, um vorher eine möglichst unvoreingenommene Einstellung zum Thema Digitalisate zu erhalten, sollten diese nicht als Forschungsdaten assoziiert werden. Daran schließt die Frage an, ob sich die Interviewten eine Verwaltung von Forschungsdaten mit Bezug zu Digitalisaten durch die entsprechende Institution wünschen (N-4.3.), eine Frage, die auch an die Interviewpartner aus Gedächtniseinrichtungen gestellt wird. Dabei soll auch erläutert werden, was gegen andere Repositorien sprechen würde, ob die Interviewten dies als eine Aufgabe der Gedächtniseinrichtungen ansehen und ob sie selber ihre Forschungsdaten in ein solches Repository zu den entsprechenden Quellen legen würden. Zum Abschluss sollen die Interviewpartner kurz skizzieren, wie sie sich als Nutzer das ideale digitale Angebot einer Gedächtniseinrichtung vorstellen würden (N-4.4.), um daraus auch Rückschlüsse für Gedächtniseinrichtungen zu ziehen.

5.4.2. Auswertung der Interviews - Nutzer.

5.4.2.1. System (N-1.).

Beim Einstieg in ein System zeigt sich unter den interviewten Nutzern eine klare Präferenz. Sowohl Nutzer 2 als auch Dr. Kohring bevorzugen beide die Suche. Für Nutzer 2 ermöglicht diese das schnelle Finden von Themen, Zeitepochen oder Personen auf Grundlage vorheriger Informationen, für Dr. Kohring ist die Suche elementar, um im Alltag sehr schnell sehr große Textkorpora durchsuchen zu können. Hier kommt eine der Besonderheiten der Antikenforschung zum Tragen, da die vorhandenen Texte digital bereits gut mit einem Thesaurus der griechischen und lateinischen Sprache erschlossen sind und daher mit einer Standard-Antik-Tastatur schnell und ohne weitere Hilfe von

Metadaten oder Übersetzungen durchsucht werden können. Für beide ist die Facettierung ein Nice-to-have, gerade wenn sie zur Suche passt, aber kein Muss.

Nutzer 1 verwendet dagegen beide Formen der Suche, da mit beidem immer wieder gute Ergebnisse erreicht werden können.

Auch zu Portalen wie der DDB oder fachspezifischen Angeboten wie Clio oder DigiZeitschriften findet sich eine klare Einstellung. Für die eigene Forschung werden diese nämlich kaum genutzt, da diese eher ablenkend sind. Als erfahrene Forscher wissen die Interviewten, wo sie nach was zu suchen haben. Dr. Kohring beschreibt hier für den Fachbereich Alte Geschichte z.B. die Linksammlung von Udo Hartmann, welche am Geschichtsinstitut der Humboldt-Universität gehostet wird oder für Überblicke, auf Grund des sehr guten hauptamtlichen Redakteurs, die deutschsprachige Wikipedia, die teilweise besser als kostenpflichtige Angebote ist. Für eine Suche nicht relevante Informationen können dagegen eher ablenkend sein. Für den Bereich der Denkmalpflege sieht Nutzer 2 in den Zusatzinformationen einen gewissen Vorteil, da es gerade in den Architektur- und Bauzeitschriften oft Querverweise zu zeitgleichen Bauprojekten gibt.

Die Portale werden dennoch von den Interviewten benutzt, nämlich um Materialien für den Unterricht und die Vorbereitung von Exkursionen zu verwenden. Die Portale, wenn sie denn gut aufbereitet sind wie z.B. das Portal zum 2000jährigen Jubiläum der Varusschlacht 2009, bieten dann zudem eine gute Lernumgebung für Studenten, welche auch gut in den Unterricht eingebunden werden kann.

Da die Interviewten überwiegend mit Text arbeiten, ist die Suche auch die wichtigste Funktion.

Nutzer 2 beschreibt zudem die Erstinformation als wichtige Funktion der digitalen Angebote, damit die dort gebotenen Zusatzinformationen in den bekannten Sachstand eingeordnet werden können.

Für Dr. Kohring bieten zusätzliche Funktionen wie z.B. die Visualisierung von Texten in Inschriften mit mehreren Darstellungsebenen eine interessante und für den Unterricht gut nutzbare Darstellungsform, ihre Anwendung in der täglichen Forschung ist jedoch eher die Ausnahme.

5.4.2.2. Anforderungen an Digitalisate (N-2.).

Zentrale Metadaten für alle interviewten Nutzer sind die klassischen bibliographischen Daten wie Titel, Autor, Ort und Zeit. Weitere Informationen sind aber immer hilfreich, da sie die Glaubwürdigkeit eines Digitalisates erhöhen und die Grundlage für weitere Arbeiten bilden. „Je detaillierter die Angaben sind, desto eher bin ich geneigt, diese Quelle zu benutzen“ (Nutzer 1, Nf-2.1.). Für die Archivarbeit bemerkt Nutzer 2, dass die Position im Hierarchiebaum des Archivs hilfreich ist, da man dann mit Hilfe der Tektonik weitere Kontextinformationen und vielleicht Querverweise erhält. Auch eine Verlinkung von Namen oder Hinweise auf weitere zeitgenössische Quellen wären ideal. Für die Alte Geschichte beschreibt Dr. Kohring, dass die benötigten Metadaten, wie sie im Thesaurus Linguae Graecae an der University of California Irvine seit den 1970er Jahren erfasst werden, vollkommen ausreichend sind. In den Textdatenbanken benötigt man Autor und

Titel, welche seit Jahrhunderten normiert sind, sowie die Kapiteleinteilungen, die durch den Teubner-Standard, die Bibliotheca Teubneriana, seit dem 19. Jahrhundert normiert sind. Diese Metadaten sind bekannt und bilden die Grundlage der Datenbanken, deren Vorteil daher in einer besseren Recherche liegt, die die zeitraubende Suche am Regal erspart.

Bei den Grundanforderung an Bilddigitalisate unterscheiden sich die Nutzermeinungen etwas. Dr. Kohring beschreibt, dass für die Antikenforschung immer die Texte im Vordergrund stehen. Wenn er daher Digitalisate betrachtet, z.B. zur Rezeption antiker Texte in späteren Handschriften, so ist vor allem der Volltext, also eine OCR-Erfassung wichtig. Eine gute grafische Darstellung mag dann interessant, für die Forschung jedoch unwichtig sein. Auch Nutzer 2 beschreibt, dass für ihn der Text am wichtigsten ist, auf einer Abbildung daher alle Informationen wie z.B. Randnotizen oder kleine Autorensignaturen klar erkennbar sein müssen. Die Farbigkeit spielt dabei nur eine nebeneordnete Rolle, wenn sich daraus nicht zusätzliche Informationen ergeben, wie z.B. bei farblichen Korrekturen in Bauplänen. Gut ist es daher, wenn sich die Einrichtungen an die DFG-Richtlinien halten. Wenn dagegen zu einem abgebildeten Objekt und nicht dem Text geforscht wird, so ist für Nutzer 1 die Farbechtheit zentral. Dieser beschreibt jedoch auch, dass man nie über ein Digitalisat schreiben, sondern die Eindrücke immer vom Original beziehen sollte. Das Digitalisat dient danach nur noch als Gedächtnisstütze.

Als Historiker arbeiten die befragten Nutzer überwiegend mit Textdaten, welche geschrieben, verwaltet und durchsucht werden. Teilweise kommen auch Bilder von den Forschungsobjekten hinzu. Das Dateiformat spielt dabei keine Rolle, solange es textuell gut verarbeitet werden kann.

Auch das Metadatenformat ist den Befragten egal, solange die Metadaten komplett und möglichst umfangreich sind. Nutzer 2 beschreibt, dass er bei der Eingabe von Informationen in Datenbanken zwar einen guten Mix aus vordefinierten Auswahlmöglichkeiten und Freitextfeldern bevorzugt, welche an die jeweiligen Informationen angepasst sein muss, die dahinter liegende Technik für ihn jedoch unwichtig ist.

Dr. Kohring bemerkt hierbei, dass er die Formate bei Bedarf zwar nachschauen könnte, aber „die Provider der Informationen haben einen Trust Value, sind in der Wissenschaft so etabliert, dass ich ihre Arbeitsweise nicht nachgucken muss, ich spare Zeit“ (Kohring, N-2.2.).

Bei der Dokumentation des Digitalisierungsprojektes gibt es unterschiedliche Meinungen. So sei grundsätzlich die Dokumentation weniger wichtig, da der Inhalt für die Forschung relevant ist. Bei einigen Forschungsbereichen, oder aus Interesse an der Wissenschaftslandschaft, können jedoch auch die historischen Hintergründe des Bestandes oder das Zustandekommen der Daten und des Projektes relevant sein. Zudem könnte so auf mögliche, in vorherigen Digitalisierungsprojekten fehlende Bestände

hingewiesen werden, die nun auch digital verfügbar sind. Es würde so also ein Überblick über den Bestand des Projektes gegeben werden.

Auch bei der bevorzugten Verarbeitungsstufe finden sich unterschiedliche Aussagen. Nutzer 1 findet eine Beantwortung der Frage im Generellen schwierig, da sich die Anforderungen von Thema zu Thema unterscheiden und mal Rohdaten, mal erschlossene Volltexte besser sind.

Nutzer 2 bevorzugt dagegen klar Volltexte, nur in Ausnahmen sind Rohdaten besser geeignet und Dr. Kohring beschreibt als Ideal eine noch höhere Verarbeitungsstufe, in der nicht nur strukturierter Volltext, sondern auch eine Positionsangabe im Text (z.B. Zeile 1, Kapitel 1 aus Buch X) für die Zitation gegeben ist, was sich bei antiken Texten gut realisieren lässt. Für Bilddaten ist es zudem gut, diese in Systeme zum Vergleichen einzubinden, so dass man sich z.B. Münzen und deren Entwicklung über eine bestimmte Zeit sowie zugehörigen Informationen anzeigen lassen kann.

Die Digitalisierung von Bild- und Textdaten sollte dabei in der höchsten Qualität geschehen, Nutzer 2 stellt jedoch die Frage, ob die Nutzer immer gleich Zugang zu allen Daten haben müssen oder ob diese nicht eher ablenken könnten.

Unterschiedliche Meinungen finden sich ebenso zur Verlinkung. Nutzer 2 sieht in Verlinkungen schöne, interessante Extras, die jedoch keine Bedingung für ein gutes System sind. Für Nutzer 1 können Verlinkungen an wichtigen Knotenpunkten sinnvoll sein, müssen dazu aber mit Fachwissen vergeben werden. Zu viele oder durch starke farbliche Markierungen aufdringliche Verlinkungen lenken dagegen ab und schaden dem Angebot nur, da man hier auf zwei Ebenen lesen müsste. In solch einem Fall würde Nutzer 1 die Arbeit in diesem System abbrechen.

Anders gestaltet es sich in der Antikenforschung, da hier die Texte wie beschrieben bereits einheitlich und international normiert sind. Dr. Kohring beschreibt hier daher das Ideal einer vollverlinkten Datenbank, welche alle Texte der Antike, unabhängig ihrer Sprache, sowie die Realenzyklopädien des 19. Jahrhunderts beinhaltet und durch automatische Querverweise Verbindungen zwischen diesen herstellt. In der Praxis scheitert eine solche Idee an den Rechteinhabern der Realenzyklopädien, die Grundidee wurde jedoch bereits im Perseus Project an der Tufts University umgesetzt, indem man die antiken Texte zusammengebracht und mit einem Überblickstext zur Geschichte der Antike verlinkt sowie über Hyperlinks mit weiteren Repositorien verbunden hat.

Eine klare Meinung der Befragten, die so auch in der Literatur vertreten wird (vgl. Stauffer 2016), findet sich beim Thema Dubletten: Für Historiker gibt es keine Dubletten. Von allen Interviewten wird daher erwartet, dass grundsätzlich alle Exemplare eines Objektes digitalisiert werden sollten. Zwar werden auch Ausnahmen genannt, vor allem bei industriell gefertigten Büchern in Bibliotheken, bei denen es keine Unterschiede gibt und die Materialität keine Rolle spielt, doch bereits bei Drucken des 18. oder 19. Jahrhunderts wird darauf hingewiesen, dass diese sich trotz gleichem Drucker oftmals in Vorsatzblatt, Reihenfolge der Bindung oder durch aktualisierte Subskriptionslisten unterscheiden können. Auch bei Kunstdrucken wie Kupferstichen sollten alle Objekte digitalisiert werden,

da diese Aussagen zum künstlerischen Prozess oder dem Marktvertrieb enthalten können. Dies gilt auch für die Alte Geschichte, wo bspw. in einer epigraphischen Datenbank an Hand von Dachziegeln mit dem Stempel der zehnten Legion, welche über 280 Jahre im gesamten römischen Imperium eingesetzt war, durch eine geographische Zuordnung die Dislokation der Truppen verortet werden kann, während ein einzelner Dachziegel allein kaum Aussagekraft hätte.

Auch zu der Frage, ob Metadaten einen gewissen Ausgleich zu Informationen bieten können, die durch die Digitalisierung verloren gehen, findet sich eine klare Antwort: Metadaten und Digitalisate können für die historische Forschung nie die Arbeit an einem Original ersetzen. Zwar helfen umfangreiche Metadaten dabei, Informationen über ein Objekt zu erlangen, für spezielle Forschungsfragen können sie auch das zentrale Element bilden, aber „Metadaten sind nur eine Hilfskonstruktion. Bis zu einer gewissen Grenze lässt sich damit arbeiten, aber wie soll man einige objekttypische Informationen ersetzen?“ (Nutzer 2, N-2.7.). Nutzer 1 beschreibt hierzu, dass gerade durch die sensomotorische Wahrnehmung Vergleiche von Objekten möglich sind, welche zu neuen Assoziationen und neuen Forschungsfragen führen, man also einen besseren Eindruck erhält, als jedes Digitalisat vermitteln könnte. Dr. Kohring beschreibt, dass die Technik irgendwann vielleicht die Mittel entwickelt, um eben solche Wahrnehmungen auch abbilden zu können, doch bis dahin bleibt die Autopsie ein Kernbestand der historischen Forschung.

5.4.2.3. Verfügbarkeit und Nachnutzung (N-3.).

Laut der reinen Wissenschaftslehre darf die Verfügbarkeit einer Quelle nicht die Forschung beeinflussen. Nutzer 1 und Nutzer 2 stimmen dem zu, für sie hat die digitale Verfügbarkeit einer Quelle keinen Einfluss auf die Quellenauswahl, sie sehen diese eher als Ergänzungen, auf die man jedoch nicht angewiesen ist. Dr. Kohring schränkt dabei ein, dass die Theorie das so vorsieht, es in der Praxis jedoch eine rationale Entscheidung ist, sich wenn vorhanden auf vertrauenswürdige digitale Quellen zu stützen, da dies einfach Zeit spart.

Vertrauenswürdige Digitalisate heißt dabei vor allem aus Gedächtniseinrichtungen. Deren Digitalisierungsarbeit ist inzwischen anerkannt, zudem können die Quellen in der Regel kontrolliert werden, was jedoch kein Misstrauen, sondern vor allem einen Schutz vor Verarbeitungsfehlern in der Digitalisierung ist. Das ist besonders bedeutsam, wenn die Quelle die Grundlage für ein Hauptargument einer Publikation liefert. Dies ist jedoch ein Problem des Umgangs mit Quellen und Editionen, die mediale Frage, ob digital oder analog, ist dabei unerheblich.

Generell gilt, dass sowohl Zeit, Geld und Personal, welches in die Erstellung der Digitalisate gesteckt wurde, das Vertrauen in diese stärken. Hinzu kommt die Reputation der Einrichtung oder, bei Projekten, der jeweiligen fachlichen Bearbeiter.

Digitalisaten aus privaten Projekten oder von Laienforschern, wie sie z.B. in der Denkmalpflege vorkommen, wird dagegen mit hoher Skepsis begegnet, diese werden genau kontrolliert.

Der Umgang mit den eigenen Forschungsdaten entspricht jeweils persönlichen Einschränkungen, Verpflichtungen und Vorzügen.

So beschreibt Nutzer 1, dass ein Großteil der Arbeit, die Einarbeitung, Konzeption, Strukturierung etc. wie beschrieben aus gesundheitlichen Gründen analog erfolgen muss, hier stehen dementsprechend Bücher, Exzerpte und Forschungstopographien im Arbeitsraum im Vordergrund. Erst für die tiefere Einarbeitung und Ergänzungen kommen digitale Quellen hinzu. Gedächtniseinrichtungen können bei dieser Arbeitsweise kaum unterstützen.

Nutzer 2 beschreibt, dass er eher lokal arbeitet, nur zum Austausch großer Daten wie Bilder oder Pläne werden auch Cloud-Dienste genutzt. Wichtige Daten wie Berichte oder Texte für eine Publikation würde er hier jedoch nicht hochladen. Allgemein hat er sich beim Forschungsdatenmanagement aber an die Regeln der Geldgeber zu halten, ist nach Projekt- und Vertragsende damit aber auch nicht mehr für ihre Zugänglichkeit verantwortlich. Zudem erklärt er, dass die Daten, nachdem das Ergebnis publiziert worden ist, nicht mehr nötig sind, da für weitere Forschungen ja die Publikation vorhanden ist. Damit beschreibt er die klassische Sicht der Geisteswissenschaften: Die Monographie ist als Ergebnis veröffentlicht, Daten aus dem Arbeitsprozess sind damit veraltet.

Direkt oder indirekt hat man sich zudem nach den Anforderungen der Fachcommunity auszurichten. Wenn hier jeder z.B. Cloud-Dienste verwendet, so ist man praktisch gezwungen, diese selber zu benutzen.

Auch Dr. Kohring arbeitet lieber lokal als mit Cloud-Diensten, was jedoch praktische Gründe hat. Da Festplattenspeicher heute kompakt ist und praktisch nichts mehr kostet, ist dieser immer verfügbar. Man kann also mehrere Terabyte an Speicher herumtragen, an fast jeden PC anschließen und ist so unabhängig vom Netz arbeitsfähig. Er beschreibt dies jedoch als eine persönliche Einstellung, viele, gerade jüngere, Kollegen arbeiten dagegen nur noch cloud-basiert.

Obwohl alle interviewten Nutzer also lokal arbeiten, wird die Downloadfunktion überwiegend nur als schöne, teilweise auch sinnvolle, aber nicht nötige Arbeitserleichterung gesehen, da der Aufwand für Bereitstellung und Rechteeinholung bekannt ist. Nur Nutzer 1 sieht hierin eine wirkliche Arbeitserleichterung, da man so bspw. Texte, auf die in Fußnoten verwiesen wird, sofort nachprüfen kann, um ihre Relevanz für die eigene Arbeit und die Plausibilität des Argumentes zu kontrollieren.

Die Downloadfunktion wird zwar als positiv bewertet, Schnittstellen zum Harvesten großer Datenmengen werden dagegen nicht benutzt.

Dr. Kohring begründet das für seinen Bereich damit, dass die Quellenbestände aus der Antike einfach begrenzt, also endlich sind und man diese begrenzten Bestände daher auch alle einmal gesehen haben will. Es gibt zwar große Datenmengen in den Datenbanken, doch normale Suchanfragen grenzen die Bestände bereits so gut ein, dass die Verfeinerung der Suchanfrage oder das Schreiben von Befehlen für eine automatische Auswertung genauso lange dauern würden wie eine Auswertung der Daten per Hand und

daher sinnlos sind. Er weiß aber, dass Zeithistoriker Harvesting durchaus nutzen, für die Antike fehlen aber die Bestände in einer sinnvoll passenden Größenordnung.

Obwohl Open Access eigentlich den Nutzern zu gute kommen sollte, finden sich durchaus gemischte Ansichten zu Open Access.

Die Idee an sich wird dabei durchaus positiv aufgenommen. Open Access ist dabei jedoch keine Methode des Sparens, so Dr. Kohring, sondern eine andere Wissenschaftsidee als die bisherige, ähnlich wie Wikipedia. Durch das internationale Zusammentragen von freiem Wissen, ohne rechtliche Einschränkungen und Interessen, könnten, wie oben beschrieben, voll vernetzte, automatische Systeme entstehen, welche den Zugang zu Wissen erleichtern und neues Wissen generieren helfen.

„Eigentlich steckt da Demokratie hinter, nicht Sparen, nicht Klauen oder Plagiate. Schwarmintelligenz und Demokratie gehen zusammen und bilden dabei das Ideal der Wissenschaft im Menschheitsauftrag. Das ist die Idee, die eigentlich dahinter steckt. Und in der besten aller Welten, in der wir nicht leben, aber die ich mir vorstellen kann, wäre das natürlich super. Da ist das alles von den Steuerzahlern irgendwann mal irgendwie bezahlt worden und sollte daher für alle zukünftigen Steuerzahler verfügbar und nicht durch verschiedene Verlagsrechte eingeschränkt sein“ (Kohring, N-3.6.).

Auch Nutzer 1 sieht in Open Access eine sehr hilfreiche und sinnvolle Idee.

Kritisch wird dagegen die Umsetzung gesehen, „die Utopie des Open Access sehe ich vielfach als gescheitert an“ (Nutzer 1, N-3.6.). Der Grund dafür liegt in der zunehmenden Gewinngenerierung durch Open Access Projekte. Nutzer 1 lehnt es daher ab, wenn Gedächtnisinstitutionen zur Finanzierung von Open Access in Abhängigkeit von Unternehmen geraten, welche so gewonnene Daten anschließend vermarkten. Bei Kenntnis von solchen Abhängigkeiten würde Nutzer 1 entsprechende Digitalisate daher nicht nutzen.

Gedächtniseinrichtungen müssen als Quellen frei und öffentlich sowie finanziell unabhängig bleiben. Dies ist die Grundvoraussetzung. Finanzielle Entschädigungen für dabei entstehende Bearbeitungskosten sind daher angemessen.

Auch Nutzer 2 lehnt einen vollen Zugang zu Digitalisaten für alle Nutzer ab, er fürchtet hier einen „Duplizierungswahn ohne Ende“ (Nutzer 2, N-3.6.). Besser wäre eine Abstufung, nur JPEGs frei, TIFFs oder andere hochqualitative Dateien jedoch nur gegen Bezahlung zur Verfügung zu stellen, auch, da die Digitalisierung eine riesige finanzielle Leistung darstellt. Im Endeffekt sieht er hier aber eine Abwägung des Aufwandes und der finanziellen Leistungen auf Seiten der Institutionen.

5.4.2.4. Vorstellungen und Wünsche (N-4.).

Eine Einbeziehung in Entscheidungsprozesse der Digitalisierung in Gedächtniseinrichtungen wird von keinem der interviewten Nutzer gewünscht. Nutzer 1 sieht es vielleicht als sinnvoll an, wenn die entsprechenden Nutzer interessiert und umfassend informiert sind, also auch begründete Argumente liefern können, ansonsten ist

eine Quasi-Demokratisierung, welche Entscheidungen an Nutzer ohne genügende Informationssättigung weiterleitet, sehr schwierig. Persönlich lehnen die Befragten eine Mitentscheidung ab, da diese Aufwand und Zeit kostet. „Ich möchte nicht mit einbezogen werden. Ich will, dass die Gedächtnisinstitution diese Entscheidungen trifft und im Sinne einer staatlichen Institution löst“ (Nutzer 1, N-4.1.).

Auch bei der Frage nach den Forschungsdaten der Geschichtswissenschaft antworten alle Befragten einheitlich und entsprechen dabei dem klassischen Bild der Historiker: Alles kann untersucht werden, daher kann auch alles ein Forschungsdatum darstellen.

„Alles, was dazu gehört, dass am Ende ein begründeter, nachvollziehbar argumentierter wissenschaftlicher Text steht [und] alles was dazu gehört, diesen Text zu schreiben und ihn in der, von der wissenschaftlichen Gemeinschaft akzeptierten, nachweisbaren Form zu veröffentlichen“ (Nutzer 1, N-4.2.) sind Forschungsdaten, was auch die Form, die Art und Weise, wie der Text geschrieben ist sowie den Entstehungsprozess und Entscheidungen mit einschließt.

Dr. Kohring bemerkt hierbei, dass all diese Daten auch idealerweise nach Projektende noch nachnutzbar sein sollten, weist aber auch auf Persönlichkeitsschutzrechte, z.B. bei der Auswertung von Unternehmensarchiven, hin.

Nutzer 2 schränkt die Forschungsdaten insofern ein, als dass diese für ihn wirklich historisch sein müssen, also eine gewisse Zeit seit der Entstehung vergangen sein muss. Grundsätzlich kann jedoch alles, wenn es nur ein gewisses Alter erreicht hat, ein Forschungsdatum darstellen. Als Beispiel nennt er hierbei Exzerpte zu einigen Briefen Schinkels aus der Zeit vor dem Zweiten Weltkrieg. Diese stellen nicht nur an sich bereits eine Quelle dar, wie die Briefe zu dieser Zeit ausgewertet wurden, da die Originalbriefe im Weltkrieg vernichtet wurden, bilden die Exzerpte die einzigen noch existierenden Hinweise auf deren Inhalte und können somit in doppelter Hinsicht interessant für die Forschung sein. Auch Entscheidungen sind, gerade wenn sie sich als falsch herausgestellt haben, interessant, da sich so z.B. neue Erkenntnisse ergeben können, die durch frühere Entscheidungen nicht möglich waren.

Zum Thema der Verwaltung von Forschungsdaten durch die Einrichtung, aus welcher die Digitalisate als Grundlagen stammen, finden sich ganz unterschiedliche Einstellungen.

Für Nutzer 1 ist es selbstverständlich, dass die Gedächtniseinrichtungen solche Daten aufbewahren und auch aktiv sammeln. Dieser würde eigene Daten daher auch an eine Gedächtniseinrichtung abgeben und dort online stellen lassen, jedoch mit der Einschränkung, dass die Ergebnisse aus diesen Daten zuerst publiziert worden sind. Zuständig für diesen Kontakt zu den Forschern sind die Fachreferenten oder Sammlungsleiter, da diese die Schnittstelle zwischen Nutzern und Objekten bilden. Personaleinsparungen an dieser Stelle sind daher hochproblematisch. „Man benötigt immer auch das persönliche Gespräch, um Daten, die man selber erarbeitet hat, zu verschenken und nichts anderes tun Forscher, wenn sie ihre Daten verfügbar machen. Das ist ein Aushandlungsprozess“ (Nutzer 1, N-4.3.).

Auch für Dr. Kohring ist klar, dass es eine Einrichtung geben muss, die solche Forschungsdaten in Beziehung zu den Digitalisaten betreut, er stellt jedoch die Frage, ob

es eine Gedächtniseinrichtung sein muss. Statt dessen sollte es eine nationale oder besser supranationale Einrichtung geben, welche solche Daten sammelt und einheitlich zur Verfügung stellt, um unterschiedliche Standards und Rechtslagen zu vermeiden. Die Etablierung einer solchen Einrichtung ist aber Aufgabe der Politik, nicht von Gedächtniseinrichtungen. Er selber würde seine Daten wenn möglich verfügbar machen und hat dies auch schon getan, bei Archivakten etc. fehlen den einzelnen Historikern hierzu aber die Rechte.

Nutzer 2 sieht das Sammeln von solchen Forschungsdaten dagegen nicht als Aufgabe der Gedächtniseinrichtungen, in denen die Digitalisate liegen, sondern der Zeitschriften, in denen die Ergebnisse publiziert werden, da man als Forscher eher hier nachschauen würde. Zentral ist für ihn zudem, dass die wissenschaftliche Leistung bspw. einer Texttranskription anerkannt, der Urheber der Daten also genannt wird. Ohne Transkription kann nicht jeder Nutzer ohne weiteres mit allen Quellen arbeiten, diese stellt also eine Zugänglichmachung dar, die auch gewürdigt werden muss. Wichtig ist es daher sicherzustellen, dass diese Leistungen, genau wie bei einer analogen Edition, bibliographisch korrekt zitiert werden können. Wenn dies gegeben und der Aufwand entsprechend vergütet ist, so würde auch Nutzer 2 entsprechende Daten an Gedächtnisinstitutionen abgeben.

Das ideale digitale Angebot wird von jedem Interviewten unterschiedlich beschrieben.

Für Nutzer 1 ist die Verfügbarkeit von klaren Ansprechpartnern wichtig. „Ich möchte gerne mit kompetentem Personal über das digitale Angebot sprechen können“ (Nutzer 1, N-4.4.). Chats oder gar automatische Antwortsysteme werden abgelehnt, da die Einarbeitung hier zu viel Zeit kostet. Wichtig ist daher, dass digitale Systeme nicht zu Personalkürzungen in den entsprechenden Einrichtungen führen.

Für Nutzer 2 ist dagegen der unkomplizierte Zugang zu Daten wichtig, wobei die Systeme die Waage zwischen einer Datenflut und fehlenden Informationen halten müssen. Als Ideal beschreibt er hierzu den Architektenkatalog des Architekturmuseums der Technischen Universität Berlin, welcher herunterladbare Basisinformationen beinhaltet und zu einer Erstinformation ausreicht. Interne Evaluationen sind daher wichtig. Als weiteren Aspekt nennt er einen Abbau von Sprachbarrieren. Mindestens Navigation und Funktionen sollten bei den Angeboten großer Gedächtniseinrichtungen zwei- oder mehrsprachig, also in der Landessprache und Englisch, vorhanden sein, idealerweise auch in den für das jeweilige Thema wichtigen Forschungssprachen. Dies würde eine leichtere Zugänglichkeit gewähren und den Wissensvorsprung der Forscher mit den jeweiligen Sprachkenntnissen abbauen.

Dr. Kohring beschrieb das für ihn ideale digitale System für Gedächtniseinrichtungen bereits bei verschiedenen anderen Interviewfragen: Ein System, welches alle verfügbaren Datenbestände, ohne rechtliche Einschränkungen und bei voller Zugänglichkeit, zur Verfügung stellt und untereinander vernetzt.

Insgesamt sollte ein System also, wie auch in der Literatur beschrieben (vgl. Münster et al. 2018, S. 372f.) möglichst einfach zugänglich und intuitiv sein, keine zu hohen technischen Anforderungen haben und durch entsprechendes Personal der Gedächtniseinrichtung betreut werden. Das System sollte dabei in der Präsentation nicht

überfüllt werden, für Nutzer, die vertieft mit dem System arbeiten wollen, müssen aber alle Informationen verfügbar und leicht zugänglich sein.

6. Abgleich: Was bieten Einrichtungen, was erwarten Nutzer?

6.1. Anforderungen der Nutzer.

Trotz unterschiedlicher Präferenzen und Arbeitsweisen sind gewisse Gemeinsamkeiten bei den interviewten Nutzern zu erkennen, welche sich als Anforderungen formulieren und mit den Praktiken und Möglichkeiten der Gedächtniseinrichtungen abgleichen lassen.

Bei der Verwendung des Systems bevorzugen die Nutzer in der wissenschaftlichen Praxis die Suchfunktion, durch Browsen und Facettierung können weitere gute Ergebnisse erzielt werden. Gedächtniseinrichtungen sollten in ihren Systemen daher mindestens eine Suchfunktion, besser aber auch verschiedene andere Sucheinstiege unterstützen, um möglichst vielen Nutzern einen einfachen Zugang zu ermöglichen. Wie aus den Evaluationen der Websites hervorgeht, werden diese Zugänge auch von allen untersuchten Einrichtungen angeboten.

Weitere Funktionen im digitalen System, wie sie mehr oder weniger von den Institutionen angeboten werden, stellen für die interviewten Nutzer zwar schöne Zusatzmöglichkeiten dar, dürfen die Kernaufgabe der digitalen Angebote, die Suche, jedoch nicht beeinträchtigen.

Portale werden von den Forschern dagegen kaum genutzt, da diese eher ablenkend sind. Zwar finden sich hier manchmal Querverweise, dies sind jedoch nur Zusatzinformationen. Dennoch sollten Gedächtnisinstitutionen Portale bespielen, da sie gerade für Studierende gute Überblickswerke darstellen. Besonders eignen sich hier die großen, bekannten Portale wie die DDB oder die Europeana, welche auch von allen untersuchten Einrichtungen (perspektivisch) bespielt werden. Wenn zudem bestimmte, thematische Zielgruppen angesprochen werden sollen, was aktuell jedoch in keiner der Einrichtungen im Vordergrund steht, dann sollten auch deren spezifischen Portale stärker fokussiert werden.

Metadaten bilden die Grundlage für die Arbeit mit digitalen Quellen, Gedächtniseinrichtungen sollten daher darauf achten, mindestens die bibliographischen Grunddaten, idealerweise aber auch so viele Zusatzinformationen wie möglich zu erfassen. Auch wenn sich die klassischen bibliographischen Grunddaten in allen Einrichtungen finden, werden doch die Unterschiede zwischen Museum und Bibliothek deutlich. Hier sollte eine gewissen Anpassung erfolgen, so dass objektspezifische Angaben auch in Bibliotheken, Strukturangaben (soweit das Objekt es erlaubt) auch im Museum zu finden sind. Für die Nutzer sind interne Erfassungsarten und Metadatenformate dabei unwichtig, wichtig ist es ihnen, so viele Informationen zu dem Digitalisat bzw. Objekt zu erhalten wie möglich. Das gilt für Metadaten ebenso wie für die eigentlichen Digitalisate, welche alle digitalisierbaren Informationen des Originals (Randbemerkungen, farbliche Kennzeichnungen, ...) enthalten sollten. Diese sind erfüllt, wenn bei der Digitalisierung die entsprechenden DFG-Empfehlungen eingehalten werden,

welche für alle untersuchten Institutionen die Mindeststandards darstellen. Die volle Zugänglichkeit zu den Metadaten ist bisher jedoch nur in den beiden untersuchten Bibliotheken gegeben, im Stadtmuseum sollten über die bibliographischen Grunddaten hinausgehende Metadaten, die bisher nur auf Anfrage zur Verfügung gestellt werden, möglichst auch direkt in der Sammlung Online angeboten werden.

Die befragten Nutzer arbeiten als Historiker vor allem mit Textdaten sowie teilweise mit Bildern. Dementsprechend sollten von den Objekten Bilder, wenn diese textuelle Inhalte haben auch durch OCR oder anders erschlossene Textdigitalisate, zur Verfügung stehen. Da eine regelmäßige Volltexterschließung bisher nur an der UB stattfindet, an der SBB und der SSB jedoch nur projektbezogen, sollte hier noch nachgebessert werden.

Eine ausführliche Dokumentation des Digitalisierungsprojektes wird dagegen nur von einigen Nutzern gebraucht, da für die Forschung eigentlich die Inhalte wichtig sind. Je nach Forschungsgebiet können jedoch auch historische Hintergründe zum Bestand oder Informationen zum Zustandekommen der Daten und des Projektes wichtig sein, so dass Gedächtniseinrichtungen hier diese Nutzer berücksichtigen und eine umfangreiche Dokumentation bereitstellen sollten. Dies ist aktuell noch nicht der Fall. Zwar werden die Zusammenhänge der Objekte in den Digitalisierungsprojekten in allen Einrichtungen gekennzeichnet, eine Dokumentation und auch Provenienzangaben sind jedoch nicht am Objektdatensatz, sondern auf einer eigenen Seite angegeben oder nur in Einzelfällen erfasst. Hier sollte z.B. durch Verlinkungen zwischen den Objekten und den Projektbeschreibungen eine Verbindung hergestellt werden.

Zu den Verarbeitungsstufen der Digitalisate finden sich unterschiedliche Aussagen, generell sollten Gedächtniseinrichtungen hier aber sowohl Rohdaten (also einfache Digitalisate) als auch verarbeitete Daten wie strukturierte Volltexte anbieten, so dass die Nutzer je nach Anwendungsfall wählen können. Wie aus den Webevaluationen und den Interviews hervorgeht, ist eine solche breite Bereitstellung in den beiden untersuchten Bibliotheken gegeben, in der SSB werden bisher nur die Bilder ohne Strukturdaten angezeigt, da dies technisch nicht möglich, aber in Planung, ist.

Verlinkungen können sinnvolle Erweiterungen darstellen oder im Idealfall zu großen Netzwerken über die Bestände führen, können jedoch auch stark ablenkend sein, was sich nicht nur in den Interviews (Nutzer 1, Nutzer 2), sondern auch in der Literatur zeigt, wonach sich die Methode der Geschichtswissenschaft durch Unmengen digitaler Informationen vom Finden zum Auswählen verlagert hat.⁵³ Daher sollten Gedächtniseinrichtungen Verlinkungen herstellen, diese jedoch in der Präsentation zurückhaltend oder in gesonderten Abschnitten einsetzen, so dass sie nicht von den Objekten ablenken und Nutzer, die diese nicht nutzen wollen, nicht davon gestört werden. Eine umfangreiche Verlinkung von Digitalisaten im Sinne von Linked Open Data besteht aktuell in noch keiner der untersuchten Einrichtungen und sollten daher hier umgesetzt werden.

Zum Thema der Digitalisierung von Dubletten wird von den Interviewten klar ausgesagt, dass es für Historiker keine Dubletten gibt. Auch durch kleine Unterscheidungen an den Objekten oder durch den Vergleich, z.B. in räumlicher oder zeitlicher Verteilung, von Objekten können bereits wichtige Erkenntnisse zu einem Thema gewonnen werden. Die

⁵³ Siehe Abschnitt 2.2. Forschungsdaten in der Geschichtswissenschaft.

Digitalisierung von wirklich gleichen, da industriell hergestellten, Massenwaren wie z.B. neuerer Literatur wird auch von den Nutzern als unökonomisch angesehen, doch bereits bei Büchern des 19. Jahrhunderts können sich Unterschiede finden, so dass die Gedächtnisinstitutionen im Zweifel alle Objekte digitalisieren sollten. Genau diese Vorgehensweise wird auch von den Institutionen beschrieben. Bei den industriell hergestellten Drucken wird jeweils der am besten erhaltene Band digitalisiert, außer es gibt besondere Exemplarspezifika, welche die Digitalisierung mehrerer Bände rechtfertigen. Einzigartige Objekte wie z.B. Handschriften werden digitalisiert, auch wenn mehrere ähnliche Exemplare vorliegen.

Auf die Frage, ob mit umfangreichen Metadaten ausgestattete Digitalisate Originale ersetzen können, findet sich bei allen Interviewten, sowohl von Seiten der Nutzer als auch der Institutionen, die gleiche Meinung. Gut ausgestattete Digitalisate können zwar dazu beitragen, Informationen über das Objekt zu erfahren und in einigen Spezialfällen (z.B. bei nicht zu öffnenden Büchern) den einzigen Zugang zum Inhalt oder (bei Analyseverfahren der Digital Humanities) das zentrale Element der Auswertung bilden, können für die wissenschaftliche Nutzung aber nicht das Original ersetzen. Ziel ist jedoch auch nicht das Ersetzen, sondern die Sichtbarmachung der Bestände. Hier reichen Digitalisate für die Nutzer, die nur am Inhalt interessiert sind, aus und bieten zudem einen enormen Zusatznutzen. Für Gedächtniseinrichtungen bedeutet dies, dass sie zu ihren Digitalisaten möglichst viele Metadaten bereitstellen, um die Zusatznutzen, die sich durch die Digitalisierung ergeben, möglichst weit und intensiv anwendbar zu machen. Die Originale müssen aber dennoch, besonders für die wissenschaftliche Nutzung, so lange wie restauratorisch möglich verfügbar gehalten werden.

Nur weil die Originale bevorzugt werden, schließt dies jedoch nicht die Nutzung von Digitalisaten für wissenschaftliche Zwecke aus. Im Gegensatz zur Anfangszeit der Publikation von Digitalisaten im Internet sind die Leistungen der Gedächtniseinrichtungen auch auf diesem Gebiet inzwischen anerkannt, so dass deren Angeboten auch als Quellen vertraut wird. Dies liegt jedoch auch an den Einrichtungen, welche in der Forschung anerkannt sind und daher einen Trust Value besitzen. Das beschriebene Vertrauen in Gedächtniseinrichtungen (vgl. Martin 2013, S. 14) und der Wandel bei der Akzeptanz von Digitalisaten (vgl. Schubert 2018, S. 229) können daher bestätigt werden. Umfangreiche Metadaten sowie die Investition von Geld, Zeit und Personal in die Qualitätskontrolle (auch durch unterschiedliche Mitarbeiter), wie sie von den Einrichtungen bereits durchgeführt werden, können das Vertrauen dabei noch weiter erhöhen.

Die interviewten Nutzer arbeiten alle eher bis absolut lokal, wobei zwar die Nutzung von Cloud-Diensten in der Community, nicht jedoch von weitergehenden Systemen zur Forschungsunterstützung oder zum kollaborativen Arbeiten erwähnt werden. Dementsprechend wird die Downloadfunktion von Datensätzen als hilfreiche und sinnvolle Arbeitserleichterung, also positiv bewertet. Auf Grund des angenommenen Aufwandes sei dies jedoch eine Ermessensfrage der jeweiligen Einrichtung. Bei einer Betrachtung der Websites zeigt sich, dass die SBB und die UB ihre digitalisierten Bestände standardmäßig auch zum Download als PDF (wie auch in der Literatur empfohlen, vgl. Schubert 2018, S.

230) oder in Bildformaten anbieten, der Aufwand hier also bereits im Workflow betrieben wird. Nur die SSB bietet keine Downloadfunktion ihrer Digitalisate (auch wenn sich die Objektdatensätze gut ausdrucken lassen), hier sollte, wenn rechtlich möglich, nachgebessert werden, um auch lokales Arbeiten ohne Netzzugang oder die Weiterverarbeitung in eigenen Systemen zu unterstützen.

Schnittstellen und Harvestingfunktion werden von den befragten Nutzern dagegen nicht verwendet, dies scheint jedoch eher von den Fachgebieten und den benötigten Datenmengen abzuhängen. Aus Nutzersicht können Gedächtniseinrichtungen Harvesting anbieten, es sollte jedoch nicht die oberste Priorität haben, zumal sich Angebote im Web mit dem entsprechenden Wissen auch durch Web Mining und Crawler einsammeln lassen. In den Interviews mit den Mitarbeitern der Gedächtniseinrichtungen wird jedoch deutlich, dass die Schnittstellen nicht nur für die Nutzer gedacht sind, sondern auch zum Datenaustausch zwischen Institutionen oder der Bereitstellung von Daten bspw. für die DDB/Europeana verwendet werden. Aus diesem Grund werden die Schnittstellen angeboten, auch wenn sie vielleicht weniger von Nutzern verwendet werden als andere Funktionen.

Die Grundidee des Open Access wird von den Nutzern überwiegend positiv bewertet, es werden jedoch auch der Aufwand und die Kosten in der Praxis gesehen, welche Gedächtniseinrichtungen für die Digitalisierung aufwenden müssen. Bearbeitungsgebühren für den Zugang zu qualitativen Digitalisaten werden daher als angemessen betrachtet. Wichtiger ist statt dessen, dass die Gedächtniseinrichtungen frei und unabhängig bleiben und sich zur Realisierung von Open Access für die Digitalisierung nicht an Unternehmen binden müssen, welche die Daten nachher (exklusiv) vermarkten könnten. Diese Unabhängigkeit und das Vermeiden einer ausschließlich gewinnorientierten Vermarktung von Digitalisaten sind wichtiger als ein absoluter Open Access Zugang zu allen hochwertigen Digitalisaten. Ob dabei Bearbeitungsgebühren erhoben werden, ist eine Abwägung der jeweiligen Institution. Die hier untersuchten Einrichtungen erfüllen daher diese Nutzeranforderung. Die Geldmittel für die Digitalisierung stammen größtenteils von öffentlichen Förderern wie der DFG, Bund oder Land. Dementsprechend werden die meisten Digitalisate, soweit urheberrechtlich möglich, auch frei angeboten. Die SBB setzt auf Public Domain-, die UB auf CC-BY-SA-NC-Lizenzen. Nur in der SSB sind die Digitalisate zwar sichtbar, aber nicht frei nutzbar, auch wenn aktuell der Umstieg auf CC-Lizenzen angestrebt wird. Die angebotene Qualität der Digitalisate reicht für die meisten Anwendungsfälle vollkommen aus, wird dennoch ein unverändertes Master-Digitalisat benötigt, so muss in allen Einrichtungen mindestens eine Bearbeitungsgebühr gezahlt werden. Ein Änderungsbedarf an dieser Praktik von der Nutzerseite aus besteht nicht, da die hier befragten Nutzer dies als angemessen, wenn auch nicht unbedingt ideal, ansehen.

Eine Miteinbeziehung der Nutzer in Entscheidungsprozesse der Digitalisierung wird von diesen weder erwartet noch gewünscht. Die Auswahl der Digitalisate und der Formate ist Aufgabe der Gedächtniseinrichtung als spezialisierter Institution, es wird erwartet, dass sie diese Aufgabe mit ihrem Spezialwissen erfüllt. Die Befragungen, wie sie von den untersuchten Einrichtungen in speziellen Communities für bestimmte Sammlungsbereiche

durchgeführt werden, sind daher ausreichend, eine allgemeine Miteinbeziehung der Nutzer wird von diesen nicht erwartet.

Es sollte dabei jedoch weniger von einer Auswahl als von einer Priorisierung der Objekte ausgegangen werden. In den Interviews wird nochmals deutlich, dass für Historiker jedes Objekt (zumindest ab einem bestimmten Alter) auch Forschungsgegenstand sein kann. Daher sind diese Objekte, aber auch alle damit in Zusammenhang stehenden Informationen, Metadaten oder Entscheidungen im Forschungsprozess, insgesamt also alles, was zu einem wissenschaftlichen Ergebnis beiträgt, Forschungsdaten. Für Gedächtniseinrichtungen bedeutet dies, dass sie, wie bereits oben geschildert, so viele Informationen wie möglich aufbereiten und zur Verfügung stellen sollten, da diese für bestimmte Forschungsfragen wichtig werden könnten.

Das aus Digitalisaten entstandene Forschungsdaten gesammelt und dauerhaft gesichert werden sollten, wird, solange die wissenschaftliche Leistung anerkannt und sichtbar bleibt, von allen Interviewpartnern bestätigt. Unklar ist jedoch, ob dafür die jeweilige Gedächtniseinrichtung, ähnlich wie bei analogen Belegexemplaren, zuständig ist oder ob die Daten durch eine zentrale Stelle gesammelt und verwaltet werden sollten. Die Gedächtniseinrichtungen würden ihre Daten dann mit dieser Stelle verlinken, aber kein eigenes Portal aufbauen, welches mit den Portalen anderer Einrichtungen konkurrieren würde. Wie auch in den Interviews beschrieben ist dies keine Entscheidung, die eine Institution alleine treffen kann. Statt dessen sollten hier die Verbünde bzw. die Politik tätig werden, indem bspw. die DNB mit der Verwaltung von Forschungsdaten beauftragt wird oder diese auch in der DDB eingespielt werden können. Da es aktuell jedoch keine Lösung hierzu gibt, sollten Gedächtniseinrichtungen Forschungsdaten, die ihnen angeboten werden, annehmen und entsprechend verwalten bzw. sogar verfügbar machen, bei der Etablierung eines zentralen Angebotes aber auch an dieses abgeben.

Auch wenn sich die Vorstellungen eines idealen digitalen Angebots bei den befragten Nutzern unterscheiden, lassen sich als Kernaspekte ein unkomplizierter Zugang und die Verfügbarkeit der Digitalisate, ein Ausgleich zwischen Informationsflut und Informationsmangel in der Darstellung, Verlinkungen von Objekten (auch über Einrichtungsgrenzen hinweg) sowie klare Ansprechpartner bei Fragen herausbilden. Dies deckt sich mit den in anderen Interviewfragen identifizierten Anforderungen. Der Zugang zu den Daten und auch klare Ansprechpartner sind durch die digitalen Angebote jeweils gegeben, auch wenn hier noch nicht alle verfügbaren Daten angeboten werden. Eine Verlinkung, gerade auch zu externen Angeboten, im Sinne von Linked Open Data ist dagegen wie beschrieben noch nicht gegeben und sollte daher umgesetzt werden.

6.2. Möglichkeiten der Institutionen.

Zusätzlich zu den Anforderungen, die von den Nutzern in den Interviews identifiziert werden, zeigen sich weitere Aspekte, die die Gedächtniseinrichtungen im Bezug auf die Digitalisate als Forschungsdaten leisten.

Trotz unterschiedlicher Ausstattungen (bspw. ist die SBB mit einem eigenen Digitalisierungszentrum und dem meisten Personal in der Digitalisierung von den drei untersuchten Institutionen am besten ausgestattet) und verschiedener Ausrichtungen, die sich aus dem Unterschied von inhaltsorientierten Bibliotheken und objektorientierten

Museen ergeben (vgl. Stiller 2014, S. 22), gibt es dennoch Gemeinsamkeiten in den Arbeitsweisen.

Die Digitalisierung erfolgt in der Regel vom Original, um eine möglichst hohe Qualität zu erreichen, wie sie auch von den Nutzern gefordert wird. Gesichert wird dies auch durch doppelte Qualitätskontrollen im Workflow und das Vermeiden von nachträglichen Veränderungen publizierter Digitalisate. Zertifikate für Forschungsdatenrepositorien werden dabei für die digitalen Sammlungen nicht angestrebt, die Nutzer erwarten dies jedoch auch nicht, da sie den Gedächtniseinrichtungen als etablierten Institutionen bereits vertrauen. Die Nutzeranforderungen im Hinblick auf die Qualität werden also bereits erfüllt und sollten so weitergeführt werden.

Social Media und teilweise Wikipedia werden von den befragten Gedächtniseinrichtungen bedient, dies jedoch eher im Sinne einer Öffentlichkeitsarbeit, weniger um Daten zur Verfügung zu stellen. Dementsprechend ist hier die interessierte Öffentlichkeit, nicht die Wissenschaft die Zielgruppe. Da die befragten Nutzer für ihre Forschung jedoch bereits kaum Portale nutzen, die konkret für Wissenschaft und Gedächtniseinrichtungen erstellt wurden, müssen die Aktivitäten der Institutionen auf allgemeinen Social Media Kanälen für die Bereitstellung von Digitalisaten als Forschungsdaten nicht weiter ausgebaut werden.

Bezüglich der Persistenz und Langzeitverfügbarkeit von Digitalisaten stellen die Nutzer keine besonderen Anforderungen, die Daten sollen nur verfügbar bleiben. Orientieren können sich Gedächtniseinrichtungen hier daher an den Anforderungen an Forschungsdaten, welche jedoch bereits erfüllt werden. Backup-Strategie und redundante Sicherungen stellen die Verfügbarkeit sicher, wobei die untersuchten Einrichtungen fest etablierte Institutionen sind, die höchstwahrscheinlich auch noch in mehr als den geforderten zehn Jahren bestehen werden und daher auch ihre Daten bereitstellen können. Die Aufgabe der Langzeitarchivierung wurde entweder an Dienstleister ausgelagert oder es wird aktuell eine Strategie dazu entwickelt. Durch die Verwendung von übergreifenden Standards soll zudem das Veralten von Digitalisaten verhindert werden, sollte es doch zu Formatveränderungen kommen, so lassen sich die Daten mittels Konvertierung/Migration relativ leicht anpassen. Neuaufnahmen werden aus Gründen der Effektivität und Bestandsschonung meist abgelehnt, sollte es dennoch dazu kommen, werden alte, publizierte Digitalisate in allen Einrichtungen erhalten, da diese auch Zustandsfotos des Objektes darstellen und um so die PURL/DOI, wenn vorhanden, gültig zu halten. Die Dokumentation der Forschungsdatenerstellung, also des Digitalisierungsprozesses, besteht zwar in allen Einrichtungen, sollte jedoch noch stärker mit den jeweiligen Objekten verknüpft und so auffindbar werden.

Fragt man nach möglichen Verbesserungsvorschlägen und Wünschen, so werden die bessere Zusammenarbeit mit den Nutzern sowie ein freier Umgang mit gemeinfreien Werken genannt. Dies ist interessant, da beide Aspekte von den Nutzern als nicht allzu notwendig erachtet werden: Entscheidungen sollen durch die Institutionen getroffen werden, Bearbeitungsgebühren für Master-Digitalisate sind akzeptiert. Für Gedächtnisinstitutionen bedeutet dies, dass sie sich in diesen Bereichen zwar bemühen können, z.B. indem Daten frei angeboten werden oder man bekannte Nutzergruppen

nach ihren Wünschen befragt, sie ihre Arbeitsprozesse aber nicht zu stark von einem erwarteten Feedback der Nutzer abhängig machen sollten.

Einen weiteren interessanten Aspekt bringt die Frage nach den finanziellen Mitteln hervor, auch wenn die jeweiligen Einrichtungen oder ihre Mitarbeiter hierauf wenig Einfluss haben. Bei unbegrenzten Mitteln wird in allen Institutionen eine Erhöhung der Quantität sowie weitere Möglichkeiten der internen wie externen Verknüpfung von Informationen und Objekten genannt, beides Aspekte, die besonders den Nutzerwünschen nach möglichst vielen, verlinkten Informationen entsprechen.

7. Fazit: Wie sollten Digitalisate als Forschungsdaten nutzbar gemacht werden?

Insgesamt zeigt sich, dass sich die Nutzer vor allem möglichst intuitive und einfache, nicht zu überladene digitale Systeme wünschen, die bei Bedarf jedoch so viele Informationen wie möglich bereitstellen, also ein umfangreiches Metadaten-set und eine möglichst genaue Objektbeschreibung. Spezielle Funktionen werden weniger gebraucht und in der Regel nur als schöne Zusatzmöglichkeiten aufgefasst. Positiv wird dagegen eine Verlinkung, ideal als Linked Open Data, gesehen, da die Objekte so miteinander in verschiedenen Verbindungen stehen und Querverweise hergestellt werden können, ein großer Vorteil der digitalen Verfügbarkeit. Der Zugang zu den Originalen bleibt dennoch zentrale Nutzeranforderung. Trotz dem Vorliegen von Digitalisaten werden, zumindest in den objektorientierten Forschungsdisziplinen, die Wissenschaftler auch weiterhin persönlich die Kulturerbeinstitutionen aufsuchen.

Für Gedächtniseinrichtungen bedeutet dies, dass sie so viele Informationen und Daten wie möglich bereitstellen (und idealerweise verlinken) sollten, um die unterschiedlichsten Nutzerbedürfnisse befriedigen zu können. Die Informationen sollten, mit Ausnahme der Basisdaten, nicht sofort sichtbar, aber leicht abrufbar sein, um das Angebot übersichtlich zu gestalten. Eine schnelles Finden der gesuchten Informationen, besonders der Basisinformationen, ist dabei nicht nur für die Nutzungseffizienz wichtig, sondern auch bei der Berücksichtigung von gesundheitlichen Aspekten, wie vom hier interviewten Nutzer 1 geschildert.

Technische Aspekte wie Formate oder Standards interessieren die Nutzer wenig, solange die Inhalte gut genutzt werden können. Entscheidungen sind hierbei also Aufgabe der Einrichtungen, welche sich daher, wie auch in der Literatur beschrieben (vgl. Voß 2013, S. 9), damit zu beschäftigen haben, welche Formate wie von wem wofür genutzt werden können. Das Erstellen, Beschreiben, auffindbar und zugänglich machen der Digitalisate ist ebenfalls Aufgabe der Gedächtniseinrichtung, die erfüllt sein muss, damit die Nutzer mit den Digitalisaten arbeiten können. Dieses erhöhte Informationsangebot führt für die Einrichtungen zu einem höheren Aufwand und der Organisationsleistung, alle Daten geordnet und qualitätsgesichert anbieten zu können. Diese umfangreichere Betreuung der digitalen Bestände wie der Nutzer, welche von diesen durchaus gewünscht wird, macht es nötig, dass das entsprechende Personal nicht nur für die Digitalisierung an sich, sondern auch für die weiterführenden Angebote dauerhaft bereit gestellt wird. Dazu gehört auch die Langzeitsicherung und -archivierung der Daten, welche jedoch als eine der Grundaufgaben in Kulturerbeeinrichtungen bereits durch den Workflow abgedeckt wird.

Wie beschrieben handelt es sich bei den Ergebnissen auf Grund der eher geringen Menge an Interviews um Fallbeispiele. Für fortführende Untersuchungen zum Bereich der Digitalisate als Forschungsdaten und zur Bestätigung der hier identifizierten Aspekte an Nutzeranforderungen könnten weitere Nutzerinterviews, sowohl mit Wissenschaftlern als

Hauptzielgruppe der untersuchten Einrichtungen als auch mit anderen Nutzergruppen, durchgeführt werden. Gerade mit Bezug auf die Nutzerwünsche nach möglichst vielen Informationen und einer Verlinkung zwischen den Daten ist dies interessant, da diese Aspekte auch als Möglichkeiten bei größeren finanziellen Mitteln der Einrichtungen genannt werden. Sollten sich diese Nutzeranforderungen bestätigen, hätten Gedächtniseinrichtungen ein weiteres Argument gegenüber den Geldgebern, die Digitalisierung und Aufbereitung der Daten (noch stärker) zu fördern.

Auch weitere technische und thematische Entwicklungen könnten für weitere Arbeiten berücksichtigt werden, wie z.B. die Vernetzung von Services und deren Angebot in unterschiedlichen Sprachen und Formaten (vgl. Crane et al. 2007, S. 120f.) oder das Suchen und Erschließen von digitalen Bildern, angefangen bei Crowdsourcing und Tagging für eine nutzerbezogenere Erschließung oder Empfehlungssysteme (vgl. Münster et al. 2018, S. 376 sowie S. 380) bis hin zu Möglichkeiten des Machine Learning für die semantische Beschreibung physischer Kulturerbeobjekte (vgl. Yasser et al. 2017), welche in dieser Arbeit aus Platzgründen nicht berücksichtigt wurden. Auch Aspekte des gesellschaftlichen digitalen Wandels sollten weiter betrachtet werden, z.B. die Frage, wer Daten aus Social Media Plattformen, welche ein Bild der aktuellen Gesellschaft darstellen, für die künftige Geschichtswissenschaft archiviert.

Die Rolle von Gedächtniseinrichtungen für die Wissenschaft hat sich durch die digitalen Entwicklungen insgesamt nicht geändert, auch wenn sich die Aufgaben erweitert haben. Es wird immer noch erwartet, dass sie die nötigen Informationen sammeln, erschließen, zugänglich machen, sichern und ihre Nutzer beim Auffinden dieser Informationen unterstützen bzw. ihnen das Verwenden ermöglichen, unabhängig davon, ob es sich um analoge oder digitale Informationen, Born Digitals oder Retrodigitalisate handelt.

8. Literaturverzeichnis.

8.1. Bibliographische Referenzen.

Allianz der deutschen Wissenschaftsorganisationen [Allianzinitiative] (2010): *Grundsätze zum Umgang mit Forschungsdaten*, [online]
https://www.mpg.de/198056/Grundsaeetze_Umgang_mit_Forschungsdaten__Dokument_im_Volltext_.pdf (abgerufen am 20.04.2019, 21:33).

Steuerungsgremium der Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen [Allianzinitiative] (2017): *Den digitalen Wandel in der Wissenschaft gestalten. Die Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen. Leitbild 2018 - 2022*, [online]
<http://doi.org/10.2312/allianzoa.015> (abgerufen am 11.04.2019, 00:52).

Arbeitsgruppe Forschungsdaten der Schwerpunktinitiative Digitale Information der Allianz der deutschen Wissenschaftsorganisationen [Allianzinitiative] (2018):
Forschungsdatenmanagement. Eine Handreichung, [online]
<http://doi.org/10.2312/allianzoa.029> (abgerufen am 11.04.2019, 00:49).

Andorfer, Peter (2015): *Forschen und Forschungsdaten in den Geisteswissenschaften. Zwischenbericht einer Interviewreihe*, Göttingen: GEODOC, Dokumenten- und Publikationsserver der Georg-August-Universität, (DARIAH-DE Working Papers, Nr. 10), [online] <http://nbn-resolving.de/%20urn:nbn:de:gbv:7-dariah-2015-3-8> (abgerufen am 06.02.2019, 00:20).

Baru, Chaitanya (2007): Sharing and caring of eScience data, in: *International Journal on Digital Libraries*, Jg. 7, H. 1-2, S. 113-116.

Baumann, Silke (2014): *Langzeitarchivierung innerhalb virtueller Forschungsumgebungen im Bereich Digital Humanities*, Berlin: Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft, Nr. 353), [online]
<https://doi.org/10.18452/2096> (abgerufen am 08.02.2019, 10:29).

Beucke, Daniel (2010): *Geschäftsmodelle für die digitale Langzeitarchivierung. Das Beispiel Forschungsdaten*, Berlin: Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft, Nr. 278), [online] <https://doi.org/10.18452/2025> (abgerufen am 08.02.2019, 10:31).

Bundesministerium für Bildung und Forschung [BMBF] (2018): *Bekanntmachung. Richtlinie zur Förderung von Forschungsvorhaben zur Entwicklung und Erprobung von Kurationskriterien und Qualitätsstandards von Forschungsdaten im Zuge des digitalen*

- Wandels im deutschen Wissenschaftssystem*, [online]
<https://www.bmbf.de/foerderungen/bekanntmachung-1791.html> (abgerufen am 28.06.2019, 15:22).
- Buckland, Michael (2017): *Information and Society*, Cambridge, Massachusetts: The MIT Press (The MIT Press Essential Knowledge Series).
- Crane, Gregory/Babeu, Alison/Bamman, David (2007): eScience and the humanities, in: *International Journal on Digital Libraries*, Jg. 7, H. 1-2, S. 117-122.
- Cremer, Fabian/Klaffki, Lisa/Steyer, Timo (2018): Der Chimäre auf der Spur. Forschungsdaten in den Geisteswissenschaften, in: *O-Bib. Das Offene Bibliotheksjournal*, Jg. 5, Nr. 2, S. 142-162.
- Czmiel, Alexander/Iordanidis, Martin/Janczak, Pia/Kurz, Susanne (2005): *Retrospektive Digitalisierung von Bibliotheksbeständen. Evaluierungsbericht über einen Förderschwerpunkt der DFG*, [online]
https://www.dfg.de/download/pdf/foerderung/programme/lis/retro_digitalisierung_eval_050406.pdf (abgerufen am 06.02.2019, 00:25).
- Deutscher Bibliotheksverband e.V. [DBV] (2018): *Wissenschaftliche Bibliotheken 2025*, [online] <https://www.bibliotheksverband.de/dbv/positionen.html> (abgerufen am 06.02.2019, 00:27).
- Deutsche Forschungsgemeinschaft [DFG] (2008): *Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten*, [online]
https://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf (abgerufen am 11.04.2019, 00:52).
- Deutsche Forschungsgemeinschaft [DFG] (2015): *Leitlinien zum Umgang mit Forschungsdaten*, [online]
https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf (abgerufen am 08.02.2019, 10:21).
- Deutsche Forschungsgemeinschaft [DFG] (2016): *DFG-Praxisregeln Digitalisierung*, [online] https://www.dfg.de/formulare/12_151/12_151_de.pdf (abgerufen am 11.04.2019, 01:06).
- Deutsche Forschungsgemeinschaft [DFG] (2018): *Ausschreibung Digitalisierung archivalischer Quellen. Eine Ausschreibung im Rahmen des LIS Förderprogramms „Erschließung und Digitalisierung“*, [online]
https://www.dfg.de/download/pdf/foerderung/programme/lis/ausschreibung_archivgutdigitalisierung_2018.pdf (abgerufen am 11.04.2019, 01:06).

- Deutsche Forschungsgemeinschaft [DFG] (2019a): *Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex*, [online] https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf (abgerufen am 11.07.2019, 14:47).
- Deutsche Forschungsgemeinschaft [DFG] (2019b): *Merkblatt Erschließung und Digitalisierung*, [online] https://www.dfg.de/formulare/12_15/12_15_de.pdf (abgerufen am 12.04.2019, 01:37).
- Dierkes, Jens/Curdt, Constanze (2018): Von der Idee zum Konzept. Forschungsdatenmanagement an der Universität zu Köln, in: *O-Bib. Das Offene Bibliotheksjournal*, Jg. 5, Nr. 2, S. 28-46.
- Diesterhöft, Martin (2014): Persönlichkeits- und datenschutzrechtliche Probleme der digitalen Zugangseröffnung zu analogen Inhalten durch Bibliotheken und Archive, in: Hinte, Oliver/Steinhauer, Eric W. (Hrsg.), *Die Digitale Bibliothek und ihr Recht. Ein Stiefkind der Informationsgesellschaft? Kulturwissenschaftliche Aspekte, technische Hintergründe und rechtliche Herausforderungen des digitalen kulturellen Speichergedächtnisses*, Münster: Verlagshaus Monsenstein und Vannerdat, S. 51-84.
- European Commission (2016): H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020, [online] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (abgerufen am 08.02.2019, 10:22).
- Fickers, Andreas (2014): Der ultimative Klick? Digital Humanities, Online-Archive und die Arbeit des Historikers im digitalen Zeitalter, in: *Forum für Politik, Gesellschaft und Kultur in Luxemburg*, Nr. 337, S. 25-29.
- Fritze, Christiane/Mikuteit, Johannes (2012): *Fragen finden, forschen. Mit der Digitalen Forschungsinfrastruktur für die Geisteswissenschaften*, [online] urn:nbn:de:0290-opus-13262 (abgerufen am 06.02.2019, 00:36).
- Fuhr, Norbert/Hansen, Preben/Mabe, Michael/Micsik, Andreas/Sølvberg, Ingeborg (2001): Digital Libraries. An Generic Classification and Evaluation Scheme, in: Constantopoulos, Panos/Sølvberg, Ingeborg T. (Hrsg.), *Research and Advanced Technology for Digital Libraries. 5th European Conference ECDL 2001*, Berlin: Springer, (Lecture notes in computer science, Nr. 2163), S. 187-199.
- Given, Lisa M./Willson, Rebekah (2018): *Information Technology and the Humanities Scholar. Documenting Digital Research Practices*, in: Journal of the Association for Information Science and Technology, Jg. 69, H. 6, S. 807-819.

- Haber, Peter (2012): *Zeitgeschichte und Digital Humanities*, [online]
<https://doi.org/10.14765/zzf.dok.2.269.v1> (abgerufen am 06.02.2019, 00:35).
- Hohls, Rüdiger (2018): Digital Humanities und digitale Geschichtswissenschaften, in:
 Busse, Laura/Enderle, Wilfried/Hohls, Rüdiger/Meyer, Thomas/Prellwitz,
 Jens/Schuhmann, Annette (Hrsg.), *Clio Guide. Ein Handbuch zu digitalen Ressourcen
 für die Geschichtswissenschaften*, 2. erw. und aktualisierte Aufl., (Historisches Forum,
 Nr. 23), [online] <https://doi.org/10.18452/19244> (abgerufen am 05.04.2019, 22:24), S.
 A.1-1-A.1-34.
- Horstmann, Wolfram (2018): Zur Rolle von Bibliotheken in digitalen
 Forschungsinfrastrukturen, in: Bonte, Achim/Rehnolt, Juliane (Hrsg.), *Kooperative
 Informationsinfrastrukturen als Chance und Herausforderung. Festschrift für Thomas
 Bürger zum 65. Geburtstag*, Berlin/Boston: De Gruyter Sauer, S. 93-109.
- Hochschulrektorenkonferenz [HRK] (2014): *Management von Forschungsdaten. Eine
 zentrale strategische Herausforderung für Hochschulleitungen. Empfehlung der 16.
 Mitgliederversammlung der HRK am 13. Mai 2014 in Frankfurt am Main*, [online]
https://www.hrk.de/uploads/tx_szconvention/HRK_Empfehlung_Forschungsdaten_13052014_01.pdf (abgerufen am 11.04.2019, 01:09).
- Institute of Museum and Library Services [IMLS] (2019): *Protecting America's Collections.
 Results from the Heritage Health Information Survey*, [online]
<https://www.imls.gov/sites/default/files/publications/documents/imls-hhis-report.pdf>
 (abgerufen am 11.04.2019, 01:35).
- Kaden, Ben (2018): Warum Forschungsdaten nicht publiziert werden, in: *LIBREAS.
 Library Ideas*, Nr. 33, [online] <https://libreas.eu/ausgabe33/kaden-daten/> (abgerufen am
 04.02.2019, 14:42).
- Kaden, Ben/Kleineberg, Michael (2016): *Fu-Push Dossiers. Auswertung von
 Experteninterviews im Rahmen des DFG-Projektes „Future Publications in den
 Humanities“*, [online] <https://doi.org/10.5281/zenodo.51174> (abgerufen am 29.03.2019,
 10:56).
- Kommission Zukunft der Informationsinfrastruktur [KII] (2011): *Gesamtkonzept für die
 Informationsinfrastruktur in Deutschland*, [online] [https://www.hof.uni-
 halle.de/web/dateien/KII_Gesamtkonzept_2011.pdf](https://www.hof.uni-halle.de/web/dateien/KII_Gesamtkonzept_2011.pdf) (abgerufen am 11.04.2019, 01:15).
- Kindling, Maxi/Schirmbacher, Peter/Simukovic, Elena (2013):
 Forschungsdatenmanagement an Hochschulen. Das Beispiel der Humboldt-Universität
 zu Berlin, in: *LIBREAS. Library Ideas*, Nr. 23, [online]
<https://libreas.eu/ausgabe23/07kindling/> (abgerufen am 02.04.2019, 15:01).

Kindling, Maxi/Schirmbacher, Peter (2013): „Die digitale Forschungswelt“ als Gegenstand der Forschung. Lehrstuhl Informationsmanagement, in: *Information. Wissenschaft & Praxis*, Bd. 64, H. 2-3, S. 127-136.

Klaffki, Lisa/Schmunk, Stefan/Stäcker, Thomas (2018): *Stand der Kulturgutdigitalisierung in Deutschland. Eine Analyse und Handlungsvorschläge des DARIAH-DE Stakeholdergremiums „Wissenschaftliche Sammlungen“*, Göttingen: GEODOC, Dokumenten- und Publikationsserver der Georg-August-Universität, (DARIAH-DE Working Papers, Nr. 26), [online] <http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2018-1-3> (abgerufen am 19.07.2019, 13:49).

Klein, Lauren F./Gold, Matthew K. (2016): Digital Humanities. The Expanded Field, in: Gold, Matthew K./Klein, Lauren F., *Debates in the Digital Humanities 2016*, Minneapolis: University of Minnesota Press, [online] <https://doi.org/10.5749/9781452963761> (abgerufen am 23.03.2019, 00:09).

Klimpel, Paul/Rack, Fabian/Weitzmann, John H. (2017): *Neue rechtliche Rahmenbedingungen für Digitalisierungsprojekte von Gedächtnisinstitutionen*, 4. Aufl., Berlin: digiS Berlin.

Lang, Elke/Bohne-Lang, Andreas (2019): *Praxishandbuch IT-Grundlagen für Bibliothekare*, Berlin/Boston: De Gruyter Saur.

Ludwig, Jens/Enke, Harry (Hrsg.) (2013): *Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt*, Glückstadt: Verlag Werner Hülsbusch.

Martin, Christiane Laura (2013): Wissenschaftliche Bibliotheken als Akteure im Forschungsdatenmanagement, in: *LIBREAS. Library Ideas*, Nr. 23, [online] <https://libreas.eu/ausgabe23/03martin/> (abgerufen am 02.04.2019, 15:02).

Münster, Sander/Kamposiori, Christina/Friedrichs, Kristina/Kröber, Cindy (2018): Image libraries and their scholarly use in the field of art and architectural history, in: *International Journal on Digital Libraries*, Jg. 19, H. 4, S. 367-383.

Nauta, Gerhard Jan/van den Heuvel, Wietske (2015): *Survey Report on Digitisation in European Cultural Heritage Institutions 2015*, [online] https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/ENUMERATE/deliverables/ev3-deliverable-d1.2-europeana-version1.1-public.pdf (abgerufen am 05.04.2019, 22:23).

Nentwich, Michael/König, René (2010): Peer Review 2.0. Herausforderungen und Chancen der wissenschaftlichen Qualitätskontrolle im Zeitalter der Cyber-Wissenschaft, in: Gasteiner, Martin/Haber, Peter (Hrsg.), *Digitale Arbeitstechniken für die Geistes- und Kulturwissenschaften*, Stuttgart/Wien: Böhlau UTB, S. 143-163.

- Oßwald, Achim/Scheffel, Regine/Neuroth, Heike (2012): Langzeitarchivierung von Forschungsdaten. Einführende Überlegungen, in: Neuroth, Heike/Strathmann, Stefan/Oßwald, Achim/Scheffel, Regine/Klump, Jens/Ludwig, Jens (Hrsg.), *Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme*, Boizenburg: Verlag Werner Hülsbusch, S. 13-21.
- Pempe, Wolfgang (2012): Geisteswissenschaften, in: Neuroth, Heike/Strathmann, Stefan/Oßwald, Achim/Scheffel, Regine/Klump, Jens/Ludwig, Jens (Hrsg.), *Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme*, Boizenburg: Verlag Werner Hülsbusch, S. 137-159.
- Petras, Vivien (2013): Methoden für die Evaluation von Informationssystemen, in: Umlauf, Konrad/Fühles-Ubach, Simone/Seadle, Michael (Hrsg.), *Handbuch Methoden der Bibliotheks- und Informationswissenschaft. Bibliotheks-, Benutzerforschung, Informationsanalyse*, Berlin/Boston: De Gruyter Saur, S. 368-386.
- Puhl, Johanna/Andorfer, Peter/Höckendorff, Mareike/Schmunk, Stefan/Stiller, Juliane/Thoden, Klaus (2015): *Diskussion und Definition eines Research Data LifeCycle für die digitalen Geisteswissenschaften*, Göttingen: GEODOC, Dokumenten- und Publikationsserver der Georg-August-Universität, (DARIAH-DE Working Papers, Nr. 11), [online] <http://nbn-resolving.de/%20urn:nbn:de:gbv:7-dariah-2015-4-4> (abgerufen am 06.02.2019, 19:54).
- Rapp, Andrea (2018): Digitale Forschungsinfrastrukturen für die Germanistische Mediävistik, in: Lobin, Henning/Schneider, Roman/Witt, Andreas (Hrsg.), *Digitale Infrastrukturen für die germanistische Forschung*, Berlin/Boston: De Gruyter, (Germanistische Sprachwissenschaft um 2020, Nr. 6), S. 249-268.
- Reiche, Ruth/Becker, Rainer/Bender, Michael/Munson, Mathew/Schmunk, Stefan/Schöch, Christof (2014): *Verfahren der Digital Humanities in den Geistes- und Kulturwissenschaften*, Göttingen: GEODOC, Dokumenten- und Publikationsserver der Georg-August-Universität, (DARIAH-DE Working Papers, Nr. 4), [online] <http://resolver.sub.uni-goettingen.de/purl/?dariah-2014-2> (abgerufen am 06.02.2019, 00:32).
- Rat für Informationsinfrastrukturen [RfII] (2016): *Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*, [online] <https://nbn-resolving.org/urn:nbn:de:101:1-201606229098> (abgerufen am 11.04.2019, 00:47).
- Rat für Informationsinfrastrukturen [RfII] (2017): *Datenschutz und Forschungsdaten. Aktuelle Empfehlungen*, [online] <https://nbn-resolving.org/urn:nbn:de:101:1-201703087085> (abgerufen am 11.04.2019, 00:47).

- Sahle, Patrick/Kronenwett, Simone (2013): Jenseits der Daten. Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner 'Data Center for the Humanities', in: *LIBREAS. Library Ideas*, Nr. 23, [online] <https://libreas.eu/ausgabe23/09sahle/> (abgerufen am 29.03.2019, 10:41).
- Saracevic, Tefko (2004): *Evaluation of digital libraries. An overview*, [online] https://tefkos.comminfo.rutgers.edu/DL_evaluation_Delos.pdf (abgerufen am 28.04.2019, 15:33).
- Schaßan, Torsten (2018): Digitale Quellen. Datei- und Datenformate, in: Busse, Laura/Enderle, Wilfried/Hohls, Rüdiger/Meyer, Thomas/Prellwitz, Jens/Schuhmann, Annette (Hrsg.), *Clio Guide. Ein Handbuch zu digitalen Ressourcen für die Geschichtswissenschaften*, 2. erw. und aktualisierte Aufl., (Historisches Forum, Nr. 23), [online] <https://doi.org/10.18452/19244> (abgerufen am 05.04.2019, 22:24), S. A.6-1-A.6-26.
- Schöch, Christof (2013): Big? Smart? Clean? Messy? Data in the Humanities, in: *Journal of Digital Humanities*, Jg. 2, Nr. 3, S. 2-11.
- Schubert, Irina (2018): Nutzerbedürfnisse für retrodigitalisierte Primärquellen. Eine Metastudie und eine vergleichende Studie unter Gedächtnisinstitutionen, in: *Informationswissenschaft. Theorie, Methode und Praxis*, Bd. 5, H. 2, S. 220-250.
- Schulze, Matthias/Stockmann, Ralf (2013): Open Science und Networked Science. Offenheit und Vernetzung als Leitmotive und Visionen einer digitalen Wissenschaft im 21. Jahrhundert, in: Neuroth, Heike/Lossau, Norbert/Rapp, Andrea (Hrsg.), *Evolution der Informationsinfrastruktur. Kooperation zwischen Bibliothek und Wissenschaft*, Glückstadt: Verlag Werner Hülsbusch, S. 31-38.
- Simukovic, Elena/Thiele, Raphael/Struck, Alexander/Kindling, Maxi/Schirmbacher, Peter (2014): *Was sind Ihre Forschungsdaten? Interviews mit Wissenschaftlern der Humboldt-Universität zu Berlin*, [online] <https://doi.org/10.18452/13600> (abgerufen am 02.04.2019, 14:59).
- Söllner, Konstanze (2017): Geisteswissenschaften, in: Söllner, Konstanze/Mittermaier, Bernhard (Hrsg.), *Praxishandbuch Open Access*, Berlin/Boston: De Gruyter Saur, S. 247-253.
- Stauffer, Andrew (2016): My Old Sweethearts. On Digitization and the Future of the Print Record, in: Gold, Matthew K./Klein, Lauren F., *Debates in the Digital Humanities 2016*, Minneapolis: University of Minnesota Press, [online] <https://doi.org/10.5749/9781452963761> (abgerufen am 23.03.2019, 02:15).

Steinhauer, Eric (2017): Recht, in: Jannidis, Fotis/Kohle, Hubertus/Rehbein, Malte (Hrsg.), *Digital Humanities. Eine Einführung*, Stuttgart: J.B. Metzler, S. 345-352.

Stiller, Juliane (2014): *From Curation to Collaboration. A Framework for Interactions in Cultural Heritage Information Systems*. Dissertation im Fach Bibliotheks- und Informationswissenschaft an der Humboldt-Universität zu Berlin, [online] <https://doi.org/10.18452/16944> (abgerufen am 21.04.2019, 00:55).

Ullmann, Nadine (2016): Digitalisierung von kulturellem Erbe in den USA. Ein Bericht der Librarian in Residence 2015, in: *b.i.t.online*, Jg. 19, H. 1, [online] <https://www.b-i-t-online.de/heft/2016-01-fachbeitrag-ullmann.pdf> (abgerufen am 06.02.2019, 00:28), S. 34-41.

Voß, Jakob (2013): Was sind eigentlich Daten?, in: *LIBREAS. Library Ideas*, Nr. 23, , [online] <https://libreas.eu/ausgabe23/02voss/> (abgerufen am 02.04.2019, 15:02).

Werner, Petra (2013): Qualitative Befragungen, in: Umlauf, Konrad/Fühles-Ubach, Simone/Seadle, Michael (Hrsg.), *Handbuch Methoden der Bibliotheks- und Informationswissenschaft. Bibliotheks-, Benutzerforschung, Informationsanalyse*, Berlin/Boston: De Gruyter Saur, S. 128-151.

Wissenschaftsrat [WR] (2012): *Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020*, [online] <https://www.wissenschaftsrat.de/download/archiv/2359-12.pdf> (abgerufen am 11.04.2019, 00:46).

Yasser, A. Mustafa/Clawson, Kathy/Bowerman, Chris (2017): Saving Cultural Heritage with Digital Make-Believe. Machine Learning and Digital Techniques to the Rescue, in: Lynne Hall/Tom Flint/Suzy O'Hara/Phil Turner (Hrsg.), *HCI 2017 Digital make-believe. Proceedings of the 31st International BCS Human Computer Interaction Conference*, [online] <http://dx.doi.org/10.14236/ewic/HCI2017.97> (abgerufen am 21.03.2019, 02:13).

8.2. Nicht-bibliographische Referenzen.

1000 Schlüsseldokumente zur deutschen Geschichte im 20. Jahrhundert, https://www.1000dokumente.de/index.html?c=1000_dokumente_de&viewmode=0&l=d (abgerufen am 20.07.2019, 14:50).

DARIAH Wiki. Zusammenstellung geisteswissenschaftlicher Quellentypen, <https://wiki.de.dariah.eu/display/publicde/7.3+Zusammenstellung+geisteswissenschaftlicher+Quellentypen> (zuletzt abgerufen am 06.06.2019, 23:02).

Deutsche Forschungsgemeinschaft. Umgang mit Forschungsdaten, https://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/nachnutzung_forschungsdaten/index.html#anker62194665 (abgerufen am 27.06.2019, 23:21).

Geheimes Staatsarchiv. Digitalisierte Archivalien, https://www.gsta.spk-berlin.de/digitalisierte_archivalien_1612.html (abgerufen am 20.07.2019, 14:47).

Historisches Datenzentrum Sachsen-Anhalt, <https://www.geschichte.uni-halle.de/struktur/hist-data/> (abgerufen am 01.07.2019, 23:39).

Historisches Datenzentrum Sachsen-Anhalt. Workflow Datenmanagement, <https://www.geschichte.uni-halle.de/struktur/hist-data/datenmanagement/#anchor2919065> (abgerufen am 01.07.2019, 23:45).

HU Berlin. Institut für Geschichtswissenschaften. Dr. Andreas Kohring, <https://www.geschichte.hu-berlin.de/de/bereiche-und-lehrstuehle/alte-geschichte/alte-geschichte/personen/kohring> (abgerufen am 26.07.2019, 00:25).

Staatsbibliothek zu Berlin. Bestandserhaltung und Digitalisierung, <https://staatsbibliothek-berlin.de/die-staatsbibliothek/abteilungen/bestandserhaltung/digitalisierung/> (abgerufen am 25.07.2019, 20:40).

Staatsbibliothek zu Berlin. Digitalisierte Sammlungen, <https://digital.staatsbibliothek-berlin.de/> (abgerufen am 20.06.2019, 23:34).

Staatsbibliothek zu Berlin. Digitalisierte Sammlungen Beta-Version, <https://digital-beta.staatsbibliothek-berlin.de/> (abgerufen am 30.07.2019, 15:38).

Staatsbibliothek zu Berlin. Porträt der Staatsbibliothek, <https://staatsbibliothek-berlin.de/die-staatsbibliothek/portraet/> (abgerufen am 25.07.2019, 20:30).

Staatsbibliothek zu Berlin. StaBiKat, <http://stabikat.de/> (abgerufen am 20.06.2019, 23:45).

Staatsbibliothek zu Berlin. Über digitalisierte Sammlungen, <https://digital.staatsbibliothek-berlin.de/ueber-digitalisierte-sammlungen> (abgerufen am 20.06.2019, 23:36).

Staatsbibliothek zu Berlin. Zahlen und Fakten, <https://staatsbibliothek-berlin.de/die-staatsbibliothek/zahlen-und-fakten/> (abgerufen am 25.07.2019, 20:30).

Stadtmuseum Berlin. Geschichte der Stiftung Stadtmuseum Berlin, <https://www.stadtmuseum.de/geschichte-der-stiftung-stadtmuseum-berlin> (abgerufen am 25.07.2019, 23:40).

Stadtmuseum Berlin. Sammlung Online, <https://sammlung-online.stadtmuseum.de/> (abgerufen am 24.05.2019, 13:45).

Stadtmuseum Berlin. Sammlungen, <https://www.stadtmuseum.de/sammlungen>
(abgerufen am 25.07.2019, 23:40).

Stadtmuseum Berlin. Unsere Museen, <https://www.stadtmuseum.de/unsere-museen>
(abgerufen am 25.07.2019, 23:40).

Universitätsbibliothek der HU Berlin. Digitale Sammlungen, <https://www.digi-hub.de/viewer/> (abgerufen am 14.06.2019, 12:30).

Universitätsbibliothek der HU Berlin. Digitale Sammlungen browsen, <https://www.digi-hub.de/viewer/browse/> (abgerufen am 25.07.2019, 23:30).

Universitätsbibliothek der HU Berlin. Geschichte der Universitätsbibliothek,
<https://www.ub.hu-berlin.de/de/ueber-uns/geschichte> (abgerufen am 25.07.2019,
23:23).

Universitätsbibliothek der HU Berlin. Historische Spezialsammlungen der Bibliothek,
<https://www.ub.hu-berlin.de/de/literatur-suchen/sammlungen/sammlungen-sonderbestaende> (abgerufen am 25.07.2019, 23:30).

Universitätsbibliothek der HU Berlin. Profil, <https://www.ub.hu-berlin.de/de/ueber-uns/profil>
(abgerufen am 25.07.2019, 23:23).

9. Anlagen.

9.1. Abkürzungsverzeichnis.

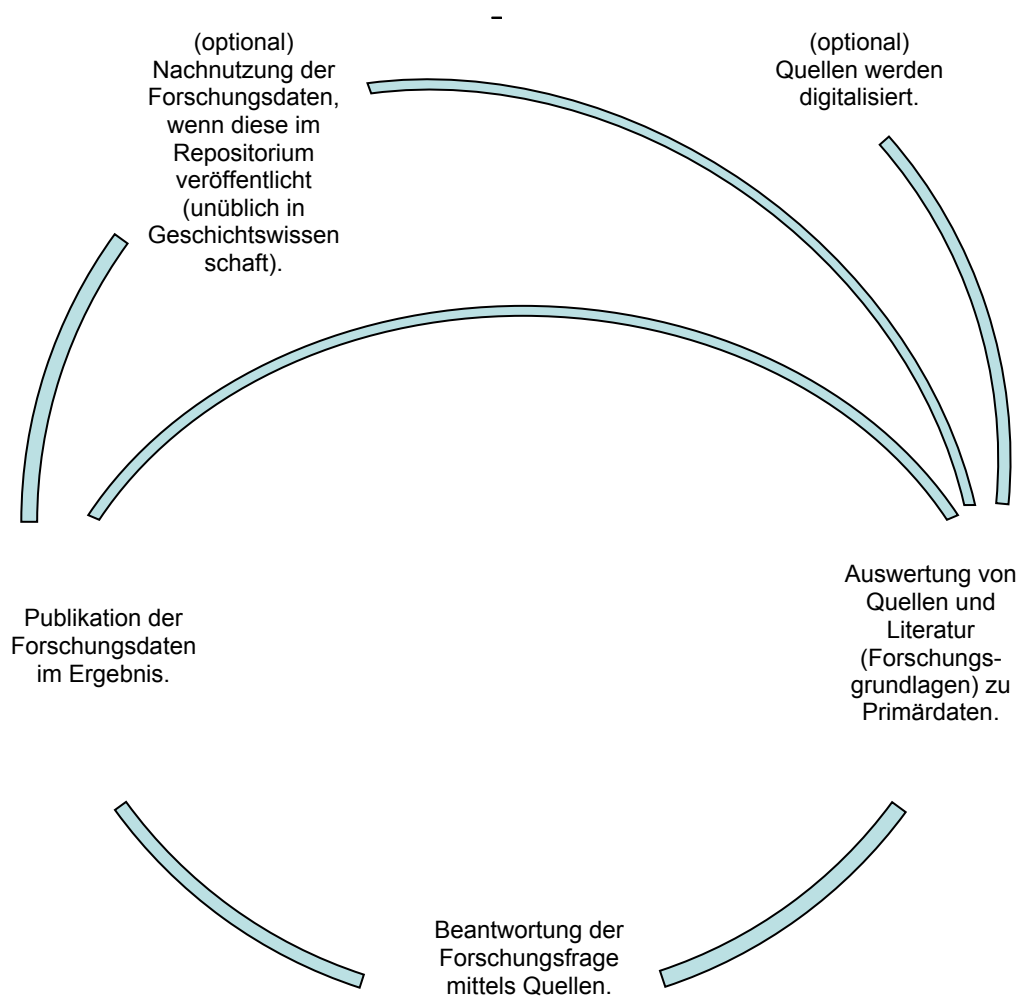
Abkürzung	Auflösung
ALTO	Analyzed Layout and Text Object
BArchG	Gesetz über die Nutzung und Sicherung von Archivgut des Bundes (Bundesarchivgesetz)
BDSG	Bundesdatenschutzgesetz
BKM	Beauftragter der Bundesregierung für Kultur und Medien
BMBF	Bundesministerium für Bildung und Forschung
DBV	Deutscher Bibliotheksverband
DDB	Deutsche Digitale Bibliothek
DFG	Deutsche Forschungsgemeinschaft
DINI	Deutsche Initiative für Netzwerkinformation
DNB	Deutsche Nationalbibliothek
DoD	Digitizing on Demand
DPI	Dots Per Inch
EAD	Encoded Archival Description
EDM	Europeana Data Model
ESE	Europeana Semantic Elements
FADGI	Federal Agencies Digital Guidelines Initiative
FID	Fachinformationsdienst
GBV	Gemeinsame Bibliotheksverbund
GG	Grundgesetz für die Bundesrepublik Deutschland
GND	Gemeinsame Normdatei
HRK	Hochschulrektorenkonferenz
HTML	Hypertext Markup Language
IIIF	International Image Interoperability Framework
IMLS	Institute for Museum and Library Services
JPEG	Joint Photographers Expert Group
JSON	JavaScript Object Notation
KII	Kommission „Zukunft der Informationsinfrastruktur“
KOBV	Kooperativer Bibliotheksverbund Berlin-Brandenburg
KunstUrhG	Gesetz betreffend das Urheberrecht an Werken der bildenden Künste und der Photographie
MARC	Machine-Readable Cataloging
METS	Metadata Encoding & Transmission Standard
MP3	MPEG-2 Audiolayer III
MPEG	Moving Picture Experts Group
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OCR	Optical Character Recognition

Abkürzung	Auflösung
OPAC	Online Public Access Catalogue
PDF	Portable Document Format
PNG	Portable Network Graphics
RDS	Resource Discovery System
Rfll	Rat für Informationsinfrastrukturen
SBB	Staatsbibliothek zu Berlin, Preußischer Kulturbesitz
SGML	Standard Generalized Markup Language
SSB	Stiftung Stadtmuseum Berlin
TEI	Text Encoding Initiative
TIFF	Tagged Image File Format
UB	Universitätsbibliothek der Humboldt-Universität zu Berlin
UrhG	Gesetz über Urheberrecht und verwandte Schutzrechte (Urheberrechtsgesetz)
WAV	Waveform Audio File
WR	Wissenschaftsrat
XML	Extensible Markup Language
ZDB	Zeitschriften Datenbank
ZVDD	Zentrale Verzeichnis Digitalisierter Drucke

9.2. Forschungskreislauf in der Geschichtswissenschaft.



9.3. Forschungsdatenlebenszyklus in der Geschichtswissenschaft.



9.4. Evaluationsfragen zur Evaluation der Websites.

Wonach richtet sich die Zusammenstellung der Objekte?	
Welche bibliographischen Angaben werden gemacht?	
Werden weitere Metadaten angeboten?	
Welche einfachen Bearbeitungsfunktionen (Vergrößern, Drehen, ...) sind möglich?	
Welche Sucheinstiege bieten sich dem	

Nutzer?	
Lassen sich Daten (über Schnittstellen) exportieren/harvesten?	
Downloadfunktion?	
Druckfunktion der jeweiligen Dokumentenansicht?	
Ansicht im DFG-Viewer möglich?	
Feedback-Möglichkeit?	

9.5. Evaluationsergebnisse der Websites.

9.5.1. Evaluation Stiftung Stadtmuseum Berlin, Sammlung Online.⁵⁴

24.05.2019

Wonach richtet sich die Zusammenstellung der Objekte?	Angeboten werden die Digitalisate des Stadtmuseums Berlin. Zwar sind einige Sammlungsbereiche stärker vertreten als andere, dennoch bietet sich ein guter Querschnitt zur Übersicht. Einige Objekte sind zudem nach Menschen, Orten, Epochen oder ihren Digitalisierungsprojekten sortiert.
Welche bibliographischen Angaben werden gemacht?	Es werden Angaben gemacht zum Sammlungsbereich und der Art des Objekts, dem Autor bzw. Künstler, dem Titel, Erscheinungsjahr und -ort, Material und Technik sowie den Maßen.
Werden weitere Metadaten angeboten?	Es besteht die Möglichkeit, wissenschaftliche Anmerkungen zu verfassen, welche jedoch nicht immer vergeben sind. Durch Gruppierungskennzeichen lassen sich zudem Angaben zum Digitalisierungsprojekt machen.
Welche einfachen Bearbeitungsfunktionen (Vergrößern, Drehen, ...) sind möglich?	Wenn hochauflösgelöste Bilder vorhanden sind lassen sich diese vergrößern. Einige spezielle Objekte lassen sich auch in einer

⁵⁴ Stadtmuseum Berlin. Sammlung Online, <https://sammlung-online.stadtmuseum.de/> (abgerufen am 24.05.2019, 13:45).

	360°-Ansicht drehen oder (bei mehrseitigen Objekten) blättern.
Welche Sucheinstiege bieten sich dem Nutzer?	<p>Als Suchmöglichkeiten bieten sich dem Nutzer eine einfache Suche über einen Suchschlitz sowie eine erweiterte Suche, welche folgende Parameter bietet:</p> <ul style="list-style-type: none"> - Sammlung - Titel - Material - Technik - Person - Oberbegriff (Art des Objektes) - Schlagworte - Datierung <p>Zum Browsen kann der Nutzer verschiedene Reiter benutzen:</p> <ul style="list-style-type: none"> - Menschen <ul style="list-style-type: none"> - Berliner Typen - Kinder - Frauen - Bürgermeister/Ratsherren - Bürger - Arme/Krange - Arbeiter - Künstler/Kreative - Persönlichkeiten - Orte <ul style="list-style-type: none"> - Stadtansichten - Verkehr - Straßen/Plätze - Gebäude - Sehenswürdigkeiten - Berliner Schloss - Rathaus - Stadtbefestigung - Ausflugsziele - Epochen <ul style="list-style-type: none"> - Berlin der Gegenwart nach 1990 - Die geteilte Stadt 1945-1990 - Groß-Berlin 1920-1945 - Berlin im Kaiserreich 1871-1920 - Die königliche Residenzstadt 1710-

	<p>1871</p> <ul style="list-style-type: none"> - Die landesherrliche Residenz <p>1440- 1710</p> <ul style="list-style-type: none"> - Die mittelalterliche Stadt 1237-1440 - Berlin vor 1237 <p>- Projekte</p> <ul style="list-style-type: none"> - 360° Objektansichten - Dokumente zum Blättern - Videos - Sammlung Heinrich Zille - Nachlass Harry Croner - Top 1000x Berlin - Nachlass Rolf Goetze - Sammlung Humboldt - Nachlass Cecil Newman <p>- Best of</p>
Lassen sich Daten (über Schnittstellen) exportieren/harvesten?	Für den Nutzer gibt es keine Schnittstellen. Die Daten lassen sich nur über ein manuelles Web Mining des Seitenquellcodes sammeln.
Downloadfunktion?	Nein.
Druckfunktion der jeweiligen Dokumentenansicht?	Es gibt keine eigene Druckfunktion der Sammlung Online, die Druckfunktion des Browsers liefert für die Objektseiten aber gute Ergebnisse.
Ansicht im DFG-Viewer möglich?	Nein.
Feedback-Möglichkeit?	Ja, für jedes Objekt gibt es die Möglichkeit, eine Anmerkung/einen Kommentar zu verfassen. Zudem gibt es Kontaktinformationen im Impressum.

9.5.2. *Evaluation Staatsbibliothek zu Berlin, Preußischer Kulturbesitz, Digitalisierte Sammlungen*.⁵⁵

20.06.2019

Wonach richtet sich die Zusammenstellung der Objekte?	Angeboten werden historische Drucke, Handschriften und Nachlässe der Staatsbibliothek zu Berlin aus verschiedenen Digitalisierungsprojekten. Schwerpunkt liegt auf deutschsprachigen Drucken des 16.-19. Jh. aller Fachrichtungen sowie Ostasiatica. Hinzu kommen abendländische Handschriften und Musikhandschriften. ⁵⁶
Welche bibliographischen Angaben werden gemacht?	Es werden umfangreiche bibliographische Angaben gemacht: <ul style="list-style-type: none"> - Titel - Autor - Ort - Verlag - Erscheinungsjahr - Umfang - Sprache - Standort - Signatur - Schlagwörter - interne Kategorie - Digitalisierungsprojekt - Digitalisierer (Staatsbibliothek zu Berlin) - Indexierungsdatum des Digitalisats - Strukturtyp - Anzahl der gescannten Seiten - PPN des StaBiKat (original und digital) - PURL - Links: <ul style="list-style-type: none"> - METS - DFG Viewer - IIF Manifest - Mirador Viewer

⁵⁵ Staatsbibliothek zu Berlin. Digitalisierte Sammlungen, <https://digital.staatsbibliothek-berlin.de/> (abgerufen am 20.06.2019, 23:34).

⁵⁶ Staatsbibliothek zu Berlin. Über digitalisierte Sammlungen, <https://digital.staatsbibliothek-berlin.de/ueber-digitalisierte-sammlungen> (abgerufen am 20.06.2019, 23:36).

Werden weitere Metadaten angeboten?	Wenn vom Standard (Public Domain) abweichend, wird die gültige Lizenz angezeigt.
Welche einfachen Bearbeitungsfunktionen (Vergrößern, Drehen, ...) sind möglich?	<ul style="list-style-type: none"> - Anzeigen und Springen in den Strukturdaten - Vergrößern (auch auf reale Größe (1:1) anzeigen) - Rotation - Bildmanipulation: <ul style="list-style-type: none"> - Helligkeit - Kontrast - Sättigung - Farbinvertierung - Bookmarking über PURL - Download <ul style="list-style-type: none"> - PDF - TIFF - JPEG - Metadaten - OCR (AltoXML) - ZIP des Bandes (TIFF, JPEG, OCR)
Welche Sucheinstiege bieten sich dem Nutzer?	<p>In den Digitalisierten Sammlungen erfolgt die Suche über einen einfachen Suchschlitz. Zudem lassen sich die durchsuchten Sammlungsbereiche eingrenzen. Im Nachhinein bieten sich die Facetten Strukturtypen und Treffer mit Volltext.</p> <p>Allerdings wird auch im regulären Bibliothekskatalog (StaBiKat) auf Digitalisate von Werken hingewiesen. Hier stehen die vollen Suchmöglichkeiten einer erweiterten Suche im klassischen OPAC zur Verfügung, es wird aber im gesamten Bestand, nicht nur der Digitalisierten Sammlung gesucht. Die Suche lässt sich jedoch auf E-Books/Online-Ressourcen eingrenzen.</p> <p>Einfache Browsingfunktion, indem man eine der Suchkategorien auswählt.</p>

Lassen sich Daten (über Schnittstellen) exportieren/harvesten?	Ein Export der Metadaten kann in METS-XML oder IIIF erfolgen.
Downloadfunktion?	Ja, Einzelseiten als TIFF oder JPEG. Das gesamte Buch als PDF oder TIFF oder JPEG als ZIP. Zudem die OCR-Datei (AltoXML).
Druckfunktion der jeweiligen Dokumentenansicht?	Nein, aber das Objekt lässt sich als PDF downloaden und drucken.
Ansicht im DFG-Viewer möglich?	Ja.
Feedback-Möglichkeit?	Nur für das gesamte Angebot, nicht für einzelne Objekte.

9.5.3. Evaluation Universitätsbibliothek der Humboldt-Universität zu Berlin, Digitale Sammlungen (digi-hub).⁵⁷

14.06.2019

Wonach richtet sich die Zusammenstellung der Objekte?	In der Digitalen Sammlung werden die Digitalisate aus den historischen Sammlungen der Universitätsbibliothek präsentiert.
Welche bibliographischen Angaben werden gemacht?	Angegeben werden: - Strukturtyp - URL - URN - Persistent Identifier - Titel - Autor - Erscheinungsort - Verlag - Erscheinungsjahr - Zusätzliche Informationen (Anmerkungen) - Signatur - zugeordnete Sammlungen

⁵⁷ Universitätsbibliothek der HU Berlin. Digitale Sammlungen, <https://www.digi-hub.de/viewer/> (abgerufen am 14.06.2019, 12:30).

	<ul style="list-style-type: none"> - zugeordnetes Wissensgebiet - teilweise Volltext
Werden weitere Metadaten angeboten?	Die Objekte lassen sich auch im OPAC anzeigen, hier werden z.B. eine weitere Beschreibung, Titelverknüpfungen oder Informationen zu Sprache, Seitenanzahl oder Format vorgehalten.
Welche einfachen Bearbeitungsfunktionen (Vergrößern, Drehen, ...) sind möglich?	<ul style="list-style-type: none"> - stufenlose Vergrößerung - Drehen um je 90° - Download <ul style="list-style-type: none"> - PDF - Metadaten in METS
Welche Sucheinstiege bieten sich dem Nutzer?	<p>Grundsätzlich wird vor allem Browsing angeboten. Der Nutzer kann sich über neue Digitalisate informieren, in den Sammlungen (z.B. den Gelehrtenbibliotheken oder den historischen Drucken), den Wissensgebieten (Disziplinen) oder den erfolgreichen Projekten suchen.</p> <p>Für die klassische Suche bietet sich ein einfacher Suchschlitz, über Facetten können der Dateityp, der Strukturtyp, beteiligte Personen sowie das Erscheinungsjahr gefiltert werden.</p> <p>Da die Bestände auch im regulären Katalog der UB übernommen werden, kann hier auch die erweiterte Suche genutzt werden.</p>
Lassen sich Daten (über Schnittstellen) exportieren/harvesten?	Für die Formate MARC XML, Dublin Core und ESE ist ein Export über OAI-PMH möglich.
Downloadfunktion?	Die Metadaten lassen sich als METS-Datei, das gesamte Werk als PDF downloaden.
Druckfunktion der jeweiligen Dokumentenansicht?	Nein, aber das Objekt lässt sich als PDF downloaden und drucken.
Ansicht im DFG-Viewer möglich?	Ja.

Feedback-Möglichkeit?	Ja allgemein, nicht konkret für einzelne Objekte. Dies müsste im Feedback angegeben werden.

9.6. Interviewfragen wie an die Interviewpartner gegeben.⁵⁸

9.6.1. Interviewfragen der Institutionen.

Ziel dieser Masterarbeit ist es zu klären, wie historische Quellen digitalisiert, verwaltet und bereit gestellt werden müssen, um den Anforderungen der Geschichtswissenschaft an ihre Forschungsdaten gerecht zu werden. Mit Hilfe von Interviews sollen dabei die Praxis und Möglichkeiten von Gedächtniseinrichtungen im Umgang mit Digitalisaten sowie die Anforderungen der Nutzer an diese Digitalisate untersucht werden.

Die Interviewfragen zu den Gedächtniseinrichtungen richten sich dabei nach dem Forschungsdatenlebenszyklus der Digitalisate (Auswahl; Digitalisierung; Dokumentation, Metadaten, Erschließungsstandards; Veröffentlichung; Interoperabilität und Schnittstellen, Portale; Speicherung, Pflege und Archivierung), hinzu kommen Fragen zur Nutzung und Organisation.

Bei den Fragen hier handelt es sich um Faktenfragen, welche vor allem zur Einordnung und Vergleichbarkeit der Gedächtniseinrichtung erfasst werden.

If-2. Digitalisierung.

If-2.1. Werden bestimmte externe Standards zur Digitalisierung in Ihrer Einrichtung angewandt (z.B. die DFG-Praxisregeln zur Digitalisierung) oder folgt man hausinternen Richtlinien?

If-2.2. Welche Standards gelten konkret für die Erzeugung von Master-Bilddigitalisaten (z.B. DPI-Anzahl, Farbtiefe, Farbraum, Format, ...).

If-2.3. Wie viele Digitalisate bieten Sie insgesamt an?

If-2.4. Wie viele Objekte werden pro Jahr ungefähr neu digitalisiert?

If-2.5. Befinden sich Born Digitals in Ihrer Sammlung?

If-2.6. Findet die Digitalisierung im Haus oder durch Dienstleister statt?

- Wenn im Haus digitalisiert wird: Wie viele Mitarbeiter sind mit der Digitalisierung beschäftigt?

If-2.7. Gibt es Maßnahmen, um Doppeldigitalisierungen von Objekten zu vermeiden?

- Finden Absprachen mit anderen Gedächtniseinrichtungen statt, um hier bereits digital vorliegende Werke nicht erneut zu digitalisieren?

If-3. Dokumentation, Metadaten, Erschließungsstandards.

If-3.1. Welche Metadaten werden bei der Digitalisierung erfasst?

⁵⁸ Die „Faktenfragen“ wurden von den anderen Fragen getrennt, um den Interviewten die Möglichkeit zu geben, diese bereits vor dem Interview zu beantworten. Da dies nicht geschah, wurden die Faktenfragen daher vor den anderen Fragen zum Beginn des Interviews gestellt.

If-3.2. Stehen alle erfassten Metadaten dem Nutzer frei zur Verfügung?

- Wenn nein: Welche nicht?

If-3.3. Nach welchen Metadatenschemata werden die Digitalisate beschrieben?

If-3.4. Nach welchen Erschließungsstandards werden die Daten beschrieben? Kommen Normdaten zum Einsatz?

If-3.5. Wird wenn möglich auch der Volltext erschlossen?

- Wenn ja, in welchem Format wird dieser gespeichert?

If-4. Veröffentlichung.

If-4.1. Unter welchen Lizenzen sind die Digitalisate zugänglich?

- Welche Lizenz wird favorisiert? Warum?

If-5. Interoperabilität und Schnittstellen, Portale.

If-5.1. Können die Metadaten über maschinenlesbare Schnittstellen geharvestet werden?

- Wenn ja, welche Schnittstellen werden angeboten?

If-5.2. Werden die Digitalisate auch in andere Angebote oder Portale (z.B. die Deutsche Digitale Bibliothek/Europeana) eingespielt?

- Wenn ja, in welche Portale spielen Sie die Daten ein? Warum gerade in diese Portale?

- Wenn nein, warum nicht?

If-7. Organisatorisches.

If-7.1. Welche Fördereinrichtungen (DFG, BMBF, EU, Land Berlin, ...) unterstützen Ihre Einrichtung?

If-7.2. Besteht die Möglichkeit, neue Funktionen in das System der digitalen Angebote zu integrieren?

- Wenn ja, wie oft geschieht dies?

If-7.3. Gibt es Erhebungen darüber, welche Nutzergruppen Ihre digitalen Angebote (am meisten) nutzen?

Ziel dieser Masterarbeit ist es zu klären, wie historische Quellen digitalisiert, verwaltet und bereit gestellt werden müssen, um den Anforderungen der Geschichtswissenschaft an

ihre Forschungsdaten gerecht zu werden. Mit Hilfe von Interviews sollen dabei die Praxis und Möglichkeiten von Gedächtniseinrichtungen im Umgang mit Digitalisaten sowie die Anforderungen der Nutzer an diese Digitalisate untersucht werden.

Die Interviewfragen zu den Gedächtniseinrichtungen richten sich dabei nach dem Forschungsdatenlebenszyklus der Digitalisate (Auswahl; Digitalisierung; Dokumentation, Metadaten, Erschließungsstandards; Veröffentlichung; Interoperabilität und Schnittstellen, Portale; Speicherung, Pflege und Archivierung), hinzu kommen Fragen zur Nutzung und Organisation.

I-1. Auswahl.

I-1.1. Wie sieht der Auswahlprozess von Objekten für Digitalisierungsprojekte aus?

- In wie fern sind Kuratoren/Sammlungsleiter und Konservatoren beteiligt?
- Welche Rolle spielen Schäden an den Objekten, z.B. besonders für oder gegen eine Digitalisierung?
- Fällt die Entscheidung vor allem durch Mittelgeber oder nach eigenen Bedürfnissen?
- In wie Weit werden die Nutzerbedürfnisse berücksichtigt bzw. die Nutzer eingebunden?

I-1.2. Wie sieht der Umgang mit Dubletten aus: Werden beide Exemplare, z.B. eines Buches, digitalisiert, wird ein beispielhaftes Objekt aus beiden Dubletten erstellt oder wird sich für eines entschieden?

I-2. Digitalisierung.

I-2.1. Erfolgt die Digitalisierung immer vom Original oder wenn vorhanden auch vom Film/Microfiche?

I-2.2. Als Qualitätsstandard werden oft die DFG-Praxisregeln zur Digitalisierung genannt. Wie relevant sind diese für die Digitalisierung in ihrer Einrichtung?

I-2.3. Welche Maßnahmen zur Qualitätssicherung werden getroffen?

I-2.4. Werden Daten im Nachhinein noch verändert?

- Wenn ja, findet eine Erfassung der Veränderungshistorie statt?

I-3. Dokumentation, Metadaten, Erschließungsstandards.

I-3.1. Welche Informationen zum Kontext der Digitalisierung (Sammlungsbereich, Provenienz, Digitalisierungsprojekt, ...) werden dokumentiert und verfügbar gemacht?

I-3.2. Durch die Digitalisierung gehen einige Informationen wie z.B. die Materialität der Quelle in der Nutzung verloren. Können umfangreiche Metadaten hier einen gewissen Ausgleich schaffen?

I-4. Veröffentlichung.

I-4.1. Werden alle digitalisierten Objekte (soweit rechtlich möglich) auch in Ihren digitalen Angeboten zur Verfügung gestellt? Sind hier alle angebotenen Digitalisate frei zugänglich?

I-4.2. Werden die Digitalisate in der höchstmöglichen Qualität (Master) in der digitalen Sammlung Ihrer Einrichtung angeboten?

- Wenn nein, warum nicht, worin unterscheiden sich die angebotenen Digitalisate vom Master?
- Wenn nein, was ist nötig, um die Master-Dateien zu erhalten?

I-4.3. Welche Rolle spielen Zertifikate für Forschungsdatenrepositorien (z.B. Data Seal of Approval oder das DINI-Zertifikat) bei Ihren digitalen Angeboten?

I-5. Interoperabilität und Schnittstellen, Portale.

I-5.1. Findet eine Verlinkung zwischen Objekten (intern wie extern) statt?

I-5.2. Werden die Digitalisate von Ihrer Seite auch auf weniger wissenschaftlichen Plattformen wie Wikipedia angeboten?

- kurze Begründung

I-6. Speicherung, Pflege und Archivierung.

I-6.1. Gibt es spezielle Strategien (z.B. regelmäßige Migration, Konvertieren, ...) zur Langzeitarchivierung der Daten?

I-6.2. Wie sieht die Backup-Strategie für die Digitalisate aus?

- In welchen Abständen werden Sicherungskopien durchgeführt?
- Werden redundante Sicherungskopien geführt?
- Werden sie in anderen Formaten als im laufenden System gesichert?

I-6.3. Lohnt sich eine langfristige Archivierung von Digitalisaten (im Bezug auf Aufwand, Finanzen, Verwaltung, ...) oder sollten diese bei Bedarf (beschädigte Daten, veraltete Formate, ...) einfach neu vom Original aufgenommen werden?

I-6.4. Sollten bei einer qualitativ besseren Neuaufnahme diese Digitalisate an Stelle der alten Daten treten oder sollte hier ein neuer Datensatz angelegt und mit dem alten verlinkt werden?

I-6.5. Werden nicht mehr genutzte oder veraltete Digitalisate (z.B. veraltetes Dateiformat) aus dem System gelöscht?

- Wenn ja, unter welchen Bedingungen genau?

I-7. Organisatorisches.

I-7.1. Liegt ein Datenmanagementplan oder ähnliches zum Umgang mit den Digitalisaten vor?

- Wenn ja, welche Aspekte des Datenmanagements werden beschrieben?
- Wenn ja, wer ist für die Umsetzung verantwortlich?

I-7.2. Wer ist Ihre Zielgruppe (Fachcommunity, interessierte Öffentlichkeit, andere Einrichtungen, ...)?

I-7.3. Werden bekannte Nutzergruppen bei Entscheidungen (z.B. was als nächstes digitalisiert wird) mit einbezogen?

- Wenn ja, wie findet die Einbeziehung statt?

I-7.4. Sollten Einrichtungen die Forschungsdaten, die aus ihren Digitalisaten hervorgehen (z.B. Korpora, Textauswertungen, ...), in Form eines Repositoriums sichern und verfügbar machen?

I-7.5. Abschluss: Welche Aspekte sollten sich Ihrer Meinung nach bei der Digitalisierung in Gedächtniseinrichtungen verbessern?

I-7.6. Abschluss: Welche Möglichkeiten (Digitalisierungstempo, Erschließungstiefe, Funktionen, ...) hätte Ihre Einrichtung bei unbegrenzten Mitteln?

9.6.2. Interviewfragen der Nutzer.

Ziel dieser Masterarbeit ist es zu klären, wie historische Quellen digitalisiert, verwaltet und bereit gestellt werden müssen, um den Anforderungen der Geschichtswissenschaft an ihre Forschungsdaten gerecht zu werden. Mit Hilfe von Interviews sollen dabei die Praxis und Möglichkeiten von Gedächtniseinrichtungen im Umgang mit Digitalisaten sowie die Anforderungen der Nutzer an diese Digitalisate untersucht werden.

Die Interviewfragen zu den Nutzern richten sich dabei nach einem schematischen Arbeitsablauf beim Suchen und Nutzen der Daten (Suche im System; Anforderungen an die Daten; Verfügbarkeit und Nachnutzung). Hinzu kommen Fragen zu den Wünschen, welche die Nutzer an die Angebote von Gedächtniseinrichtungen im Bezug auf Digitalisate haben.

Bei den Fragen hier handelt es sich um Faktenfragen, welche vor allem zur Einordnung und Vergleichbarkeit der befragten Nutzer erfasst werden.

Nf-1. System.

Nf-1.1. Ist Ihnen als Sucheinstieg die Suche oder das Browsing/Facettierung wichtiger?

Nf-2. Anforderungen.

Nf-2.1. Welche Metadaten eines Digitalisates sollten auf jeden Fall, welche idealerweise erfasst werden?

Nf-2.2. Welche Grundanforderungen stellen Sie an Bilddigitalisate (z.B. Farbigkeit, Vergrößerungsfunktion, Farbraum, Qualität, OCR...)?

Ziel dieser Masterarbeit ist es zu klären, wie historische Quellen digitalisiert, verwaltet und bereit gestellt werden müssen, um den Anforderungen der Geschichtswissenschaft an ihre Forschungsdaten gerecht zu werden. Mit Hilfe von Interviews sollen dabei die Praxis und Möglichkeiten von Gedächtniseinrichtungen im Umgang mit Digitalisaten sowie die Anforderungen der Nutzer an diese Digitalisate untersucht werden.

Die Interviewfragen zu den Nutzern richten sich dabei nach einem schematischen Arbeitsablauf beim Suchen und Nutzen der Daten (Suche im System; Anforderungen an die Daten; Verfügbarkeit und Nachnutzung). Hinzu kommen Fragen zu den Wünschen, welche die Nutzer an die Angebote von Gedächtniseinrichtungen im Bezug auf Digitalisate haben.

N-1. System.

N-1.1. Nutzen Sie eher Portale (DDB, Europeana, ...) oder direkt die digitalen Angebote von Gedächtniseinrichtungen?

- Können Sie dies kurz begründen?

N-1.2. Welche Funktionen in digitalen Angeboten von Gedächtniseinrichtungen nutzen Sie am meisten bzw. sind am wichtigsten für Sie?

N-2. Anforderungen.

N-2.1. Mit was für digitalen Daten arbeiten Sie hauptsächlich?

N-2.2. Favorisieren Sie bestimmten Metadatenformate/-schemata oder Normdaten für die Erschließung?

- Können Sie dies kurz begründen?

N-2.3. Welche Informationen erwarten Sie von einer Dokumentation zu dem Digitalisat? (z.B. Provenienz, Informationen zum Digitalisierungsprozess, Informationen zum Digitalisierungsprojekt, ...)

N-2.4. Welche Verarbeitungsstufe (Rohdaten, strukturiert, Volltext, ...) ist für Sie am sinnvollsten:

- bei Textdigitalisaten?
- bei Bilddigitalisaten?

N-2.5. Wie wichtig ist für Sie die Verlinkung von Objekten untereinander?

N-2.6. Wie sollten Gedächtniseinrichtungen mit Dubletten umgehen: Beide Exemplare digitalisieren, um mögliche Unterschiede sichtbar zu machen, oder auf der Basis eines (oder mehrerer) Exemplare ein beispielhaftes Modell erstellen?

N-2.7. Durch die Digitalisierung gehen einige Informationen wie z.B. die Materialität der Quelle in der Nutzung verloren. Können umfangreiche Metadaten hier einen gewissen Ausgleich schaffen?

N-3. Verfügbarkeit und Nachnutzung.

N-3.1. In wie fern beeinflusst die digitale Verfügbarkeit der Quelle Ihre Quellenauswahl?

N-3.2. Vertrauen Sie Digitalisaten aus Gedächtniseinrichtungen als wissenschaftlichen Quellen?

- Wenn ja: Nur unter bestimmten Bedingungen?
- Wenn nein: Warum nicht? Was könnte eine Gedächtniseinrichtung hier verbessern?

N-3.3. Wie sieht Ihr persönliches Forschungsdatenmanagement aus? Stützen Sie sich auf Angebote z.B. von virtuellen Forschungsumgebungen, nutzen Sie Cloud-Dienste oder arbeiten Sie lokal?

- Welche Dateien benötigen Sie hierfür?

- Gibt es Anforderungen aus der Fachcommunity oder von Förderern?

N-3.4. Wie wichtig ist Ihnen eine Downloadfunktion für die angebotenen Dateien? Für einige Dateien besonders?

N-3.5. Benutzen Sie wenn möglich Harvesting, um große Datenmengen abzurufen?

N-3.6. Wie beurteilen Sie Open Access für Digitalisate (sowohl bezüglich Zugänglichkeit wie Nachnutzung)?

N-4. Wünsche.

N-4.1. Würden Sie als Nutzer gerne in Entscheidungsprozesse von Gedächtniseinrichtungen zur Digitalisierung einbezogen werden?

N-4.2. Was sind für Sie die Forschungsdaten der Geschichtswissenschaft?

- Sind auch Metadaten Forschungsdaten?

N-4.3. Wünschen Sie sich die Verwaltung von Forschungsdaten (Korpora, Textauswertung, ...) mit Bezug zu Digitalisaten durch die entsprechende Einrichtung?

- Wenn ja: Was spricht gegen ein disziplinäres Repositorium?

- Sehen Sie dies als Teil der Aufgaben von Gedächtniseinrichtungen?

- Würden Sie ihre Forschungsdaten selber in einem Repositorium ablegen, welches auch die Originale beinhaltet oder eher in fachspezifischen Repositorien?

N-4.4. Abschluss: Können Sie kurz skizzieren, wie für Sie das ideale digitale Angebot einer Gedächtniseinrichtung aussehen würde?

9.7. Interviewfragen mit Erläuterungen.

9.7.1. Interviewfragen der Institutionen mit Erläuterungen.

Ziel dieser Masterarbeit ist es zu klären, wie historische Quellen digitalisiert, verwaltet und bereit gestellt werden müssen, um den Anforderungen der Geschichtswissenschaft an ihre Forschungsdaten gerecht zu werden. Mit Hilfe von Interviews sollen dabei die Praxis und Möglichkeiten von Gedächtniseinrichtungen im Umgang mit Digitalisaten sowie die Anforderungen der Nutzer an diese Digitalisate untersucht werden.

Die Interviewfragen zu den Gedächtniseinrichtungen richten sich dabei nach dem Forschungsdatenlebenszyklus der Digitalisate (Auswahl; Digitalisierung; Dokumentation, Metadaten, Erschließungsstandards; Veröffentlichung; Interoperabilität und Schnittstellen, Portale; Speicherung, Pflege und Archivierung), hinzu kommen Fragen zur Nutzung und Organisation.

Bei den Fragen hier handelt es sich um Faktenfragen, welche vor allem zur Einordnung und Vergleichbarkeit der Gedächtniseinrichtung erfasst werden.

If-1. Auswahl.

Hierzu gibt es keine Faktenfragen, der Auswahlprozess soll direkt im Interview genauer beleuchtet werden.

If-2. Digitalisierung.

If-2.1. Werden bestimmte externe Standards zur Digitalisierung in Ihrer Einrichtung angewandt (z.B. die DFG-Praxisregeln zur Digitalisierung) oder folgt man hausinternen Richtlinien?

Wenn bestimmte externe Standards verwendet werden, sollten diese in der Arbeit kurz erläutert werden, um sie vergleichen zu können.

If-2.2. Welche Standards gelten konkret für die Erzeugung von Master-Bilddigitalisaten (z.B. DPI-Anzahl, Farbtiefe, Farbraum, Format, ...).

Zur Vergleichbarkeit der Einrichtungen.

If-2.3. Wie viele Digitalisate bieten Sie insgesamt an?

Zur Vergleichbarkeit der Einrichtungen.

If-2.4. Wie viele Objekte werden pro Jahr ungefähr neu digitalisiert?

Zur Vergleichbarkeit, wie viel Aufwand in die Digitalisierung gesteckt wird.

If-2.5. Befinden sich Born Digitals in Ihrer Sammlung?

Gedächtniseinrichtungen arbeiten vorrangig mit analogen Beständen. In Zukunft werden im Zuge der Digitalisierung aber auch rein digitale Objekte wichtig werden. Durch die Frage soll geklärt werden, ob sich solche Objekte bereits in den Sammlungen befinden.

If-2.6. Findet die Digitalisierung im Haus oder durch Dienstleister statt?

- Wenn im Haus digitalisiert wird: Wie viele Mitarbeiter sind mit der Digitalisierung beschäftigt?

Zur Vergleichbarkeit, wie viel Aufwand in die Digitalisierung gesteckt wird.

If-2.7. Gibt es Maßnahmen, um Doppeldigitalisierungen von Objekten zu vermeiden?

- Finden Absprachen mit anderen Gedächtniseinrichtungen statt, um hier bereits digital vorliegende Werke nicht erneut zu digitalisieren?

Digitalisierung stellt einen Kostenfaktor dar. Auch wenn auf diesen in der Arbeit nicht genauer eingegangen wird, ist es doch interessant herauszufinden, ob Kooperationen zwischen den Einrichtungen vorliegen, um Kosten und Aufwand zu senken.

If-3. Dokumentation, Metadaten, Erschließungsstandards.

If-3.1. Welche Metadaten werden bei der Digitalisierung erfasst?

Die Metadaten sind zentral bei der Erfassung der Digitalisate in den digitalen Sammlungen. Es soll daher geklärt werden, welche Metadaten gegeben sind, da sich daraus auch Nutzungsmöglichkeiten der Digitalisate als Forschungsdaten ableiten.

If-3.2. Stehen alle erfassten Metadaten dem Nutzer frei zur Verfügung?

- Wenn nein: Welche nicht?

Die Metadaten sind wie beschrieben wichtig für die Nutzung der Digitalisate. Daher soll herausgefunden werden, ob und warum einige erfasst, aber nicht veröffentlicht werden.

If-3.3. Nach welchen Metadatenschemata werden die Digitalisate beschrieben?

Je nach Art der Gedächtniseinrichtung können verschiedene Schemata die Regel sein. Es soll daher beleuchtet werden, welche Schemata in den untersuchten Einrichtungen zum Einsatz kommen.

If-3.4. Nach welchen Erschließungsstandards werden die Daten beschrieben? Kommen Normdaten zum Einsatz?

Die Erschließung ist wichtig bei der Auffindbarkeit der Digitalisate durch die Nutzer.

If-3.5. Wird wenn möglich auch der Volltext erschlossen?

- Wenn ja, in welchem Format wird dieser gespeichert?

Volltexterschließung ist aufwändig, erleichtert den Nutzern aber den Sucheinstieg.

If-4. Veröffentlichung.

If-4.1. Unter welchen Lizenzen sind die Digitalisate zugänglich?

- Welche Lizenz wird favorisiert? Warum?

Die Lizenzen regeln die Zugänglichkeit und Nachnutzbarkeit der Digitalisate, sind damit ein zentrales Element bei der Nutzung.

If-5. Interoperabilität und Schnittstellen, Portale.

If-5.1. Können die Metadaten über maschinenlesbare Schnittstellen geharvestet werden?

- Wenn ja, welche Schnittstellen werden angeboten?

Die Geschichtswissenschaft arbeitet überwiegend qualitativ, quantitative Auswertungen kommen eher in Spezialbereichen vor. Durch die Digital Humanities könnten quantitative Methoden aber verstärkt auftreten. Es soll daher geklärt werden, ob die Einrichtungen solche Methoden bereits unterstützen.

If-5.2. Werden die Digitalisate auch in andere Angebote oder Portale (z.B. die Deutsche Digitale Bibliothek/Europeana) eingespielt?

- Wenn ja, in welche Portale spielen Sie die Daten ein? Warum gerade in diese Portale?

- Wenn nein, warum nicht?

Durch die Verfügbarkeit in Portalen soll die Sichtbarkeit und Auffindbarkeit der Objekte erhöht werden. Zudem bietet sich hier ein zentraler Sucheinstieg über verschiedene Einrichtungen. Mit dieser Frage soll untersucht werden, welche Portale von den Gedächtniseinrichtungen bespielt werden, um abgleichen zu können, ob sich diese mit den von den Nutzern verwendeten Portalen decken.

If-6. Speicherung, Pflege und Archivierung-

Fragen zum Aspekt der Langzeitarchivierung werden detaillierter im Interview geklärt, da sie wichtiger Teil der Forschungsdatenmanagements sind.

If-7. Organisatorisches.

If-7.1. Welche Fördereinrichtungen (DFG, BMBF, EU, Land Berlin, ...) unterstützen Ihre Einrichtung?

Die Fördereinrichtungen können eigene Profile und Ziele haben, welche auf die Digitalisierungsstrategien der Einrichtungen wirken können. Zudem können je nach Fördereinrichtung unterschiedliche Anforderungen an die Digitalisate bzw. Forschungsdaten bestehen.

If-7.2. Besteht die Möglichkeit, neue Funktionen in das System der digitalen Angebote zu integrieren?

- Wenn ja, wie oft geschieht dies?

Wenn sich Nutzeranforderungen ändern, können neue Funktionen wichtig sein. Wenn diese ohne Probleme in das System integriert werden können, so kann dieses über einen längeren Zeitraum aktuell und nutzbar gehalten werden, die Gefahr des Veraltens besteht also nicht so schnell.

If-7.3. Gibt es Erhebungen darüber, welche Nutzergruppen Ihre digitalen Angebote (am meisten) nutzen?

Je nach Nutzergruppe können sich die Anforderungen an die Digitalisate/Forschungsdaten unterscheiden.

Ziel dieser Masterarbeit ist es zu klären, wie historische Quellen digitalisiert, verwaltet und bereit gestellt werden müssen, um den Anforderungen der Geschichtswissenschaft an ihre Forschungsdaten gerecht zu werden. Mit Hilfe von Interviews sollen dabei die Praxis und Möglichkeiten von Gedächtniseinrichtungen im Umgang mit Digitalisaten sowie die Anforderungen der Nutzer an diese Digitalisate untersucht werden.

Die Interviewfragen zu den Gedächtniseinrichtungen richten sich dabei nach dem Forschungsdatenlebenszyklus der Digitalisate (Auswahl; Digitalisierung; Dokumentation, Metadaten, Erschließungsstandards; Veröffentlichung; Interoperabilität und Schnittstellen, Portale; Speicherung, Pflege und Archivierung), hinzu kommen Fragen zur Nutzung und Organisation.

I-1. Auswahl.

I-1.1. Wie sieht der Auswahlprozess von Objekten für Digitalisierungsprojekte aus?

- In wie fern sind Kuratoren/Sammlungsleiter und Konservatoren beteiligt?
- Welche Rolle spielen Schäden an den Objekten, z.B. besonders für oder gegen eine Digitalisierung?
- Fällt die Entscheidung vor allem durch Mittelgeber oder nach eigenen Bedürfnissen?
- In wie Weit werden die Nutzerbedürfnisse berücksichtigt bzw. die Nutzer eingebunden?

Diese Frage soll einen Einstieg in das Interview bieten. Die Befragten sollen zu dieser allgemeinen Frage erst einmal erzählen, bevor die detaillierteren Fragen gestellt werden. Inhaltlich werden durch die Frage der Auswahlprozess von Digitalisierungsprojekten sowie die beteiligten Akteure in Erfahrung gebracht.

I-1.2. Wie sieht der Umgang mit Dubletten aus: Werden beide Exemplare, z.B. eines Buches, digitalisiert, wird ein beispielhaftes Objekt aus beiden Dubletten erstellt oder wird sich für eines entschieden?

Hierdurch soll geklärt werden, ob bei der Digitalisierung das Objekt oder der Inhalt im Vordergrund steht.

I-2. Digitalisierung.

I-2.1. Erfolgt die Digitalisierung immer vom Original oder wenn vorhanden auch vom Film/Microfiche?

Schließt an die vorherige Frage an, indem nochmals das Verhältnis Objekt/Inhalt beleuchtet wird.

I-2.2. Als Qualitätsstandard werden oft die DFG-Praxisregeln zur Digitalisierung genannt. Wie relevant sind diese für die Digitalisierung in ihrer Einrichtung?

Im deutschsprachigen Raum sind die DFG-Praxisregeln zur Digitalisierung, allein schon auf Grund ihres Charakters als Standard bei DFG-geförderten Projekten, weit verbreitet. Es soll daher herausgefunden werden, wie weit sich die untersuchten Einrichtungen an diesen Quasi-Standards orientieren.

I-2.3. Welche Maßnahmen zur Qualitätssicherung werden getroffen?

Die Qualitätssicherung stellt einen wichtigen Aspekt des Forschungsdatenmanagements dar. Wenn Digitalisate als Forschungsdaten behandelt werden sollen, so sollten auch Maßnahmen zur Qualitätssicherung getroffen werden.

I-2.4. Werden Daten im Nachhinein noch verändert?

- Wenn ja, findet eine Erfassung der Veränderungshistorie statt?

Persistenz ist eine zentrale Voraussetzung für die Nachnutzung von Forschungsdaten. Wenn Daten ohne die Erfassung einer Veränderungshistorie nach Veröffentlichung abgeändert werden, so können sie nicht unbedingt als wissenschaftliche Quellen angesehen werden.

I-3. Dokumentation, Metadaten, Erschließungsstandards.

I-3.1. Welche Informationen zum Kontext der Digitalisierung (Sammlungsbereich, Provenienz, Digitalisierungsprojekt, ...) werden dokumentiert und verfügbar gemacht?

Kontextinformationen sind ein weiterer wichtiger Aspekt bei der Nachnutzung von Forschungsdaten. Die Frage soll daher beleuchten, in wie weit diese Informationen erfasst werden.

I-3.2. Durch die Digitalisierung gehen einige Informationen wie z.B. die Materialität der Quelle in der Nutzung verloren. Können umfangreiche Metadaten hier einen gewissen Ausgleich schaffen?

[Meinungsfrage] Informationen wie die Materialität gehen bei der Digitalisierung verloren, können in einigen Forschungsbereichen wie z.B. der Buchwissenschaft aber zentrale Untersuchungsgegenstände sein. Durch diese Frage soll die Meinung dazu herausgefunden werden, ob (und möglicherweise welche) Metadaten hier einen Ersatz leisten können.

I-4. Veröffentlichung.

I-4.1. Werden alle digitalisierten Objekte (soweit rechtlich möglich) auch in Ihren digitalen Angeboten zur Verfügung gestellt? Sind hier alle angebotenen Digitalisate frei zugänglich?

Durch diese Frage soll herausgefunden werden, ob es Einschränkungen bei der Zugänglichkeit zu den Digitalisaten gibt.

I-4.2. Werden die Digitalisate in der höchstmöglichen Qualität (Master) in der digitalen Sammlung Ihrer Einrichtung angeboten?

- Wenn nein, warum nicht, worin unterscheiden sich die angebotenen Digitalisate vom Master?

- Wenn nein, was ist nötig, um die Master-Dateien zu erhalten?

Durch diese Frage soll herausgefunden werden, ob es Einschränkungen bei der technischen Qualität der Digitalisate gibt.

I-4.3. Welche Rolle spielen Zertifikate für Forschungsdatenrepositorien (z.B. Data Seal of Approval oder das DINI-Zertifikat) bei Ihren digitalen Angeboten?

Forschungsdatenrepositorien können sich ihre Qualität durch Zertifikate bestätigen lassen. Wenn Digitalisate als Forschungsdaten betrachtet werden sollen, so stellen die digitalen Sammlungen von Gedächtniseinrichtungen damit eine Art von Forschungsdatenrepositorien dar. Durch die Frage soll ermittelt werden, ob sonst übliche Anforderungen an Forschungsdatenrepositorien auch für Gedächtniseinrichtungen eine Rolle spielen.

I-5. Interoperabilität und Schnittstellen, Portale.

I-5.1. Findet eine Verlinkung zwischen Objekten (intern wie extern) statt?

Eine Verlinkung von Objekten würde es den Nutzern erlauben, noch leichter Zusammenhänge zu erkennen und somit weitere Kontextinformationen liefern. Es könnten so auch zusammengehörige Objekte leichter gemeinsam ausgewertet werden.

I-5.2. Werden die Digitalisate von Ihrer Seite auch auf weniger wissenschaftlichen Plattformen wie Wikipedia angeboten?

- kurze Begründung

Portale wie Europeana werden vor allem durch Gedächtniseinrichtungen bespielt, die Informationen hier stammen also von den (fachlich qualifizierten) Mitarbeitern der Institutionen. Wikipedia dagegen zeichnet sich dadurch aus, dass allen Nutzern die Möglichkeit zur Mitarbeit gegeben ist. Im wissenschaftlichen Bereich gelten Wikipedia bzw. die dort veröffentlichten Informationen daher als nicht gesicherte Quellen, da in der Regel kein eindeutiger Autor identifiziert werden kann. Einige Einrichtungen wie das Google Arts and Culture Projekt bieten ihre Digitalisate jedoch auch auf Wikipedia an, was mit der höheren Reichweite begründet werden kann.

I-6. Speicherung, Pflege und Archivierung.

I-6.1. Gibt es spezielle Strategien (z.B. regelmäßige Migration, Konvertieren, ...) zur Langzeitarchivierung der Daten?

Die Frage soll in Erfahrung bringen, wie die Langzeitarchivierung der Digitalisate ermöglicht wird.

I-6.2. Wie sieht die Backup-Strategie für die Digitalisate aus?

- In welchen Abständen werden Sicherungskopien durchgeführt?
- Werden redundante Sicherungskopien geführt?
- Werden sie in anderen Formaten als im laufenden System gesichert?

Neben der Sicherung der Nutzbarkeit muss auch der Zugang zu den Daten gewährleistet sein, indem diese erhalten bleiben (Bitstream Preservation). Die Frage soll in Erfahrung bringen, wie die Daten praktisch erhalten bleiben.

I-6.3. Lohnt sich eine langfristige Archivierung von Digitalisaten (im Bezug auf Aufwand, Finanzen, Verwaltung, ...) oder sollten diese bei Bedarf (beschädigte Daten, veraltete Formate, ...) einfach neu vom Original aufgenommen werden?

[Meinungsfrage] Die Erhaltung von Forschungsdaten ist wichtig, da diese in der Regel einmalig sind. Digitalisate können jedoch, solange das Original noch existiert, neu erzeugt werden. Ist eine umfangreiche Langzeitarchivierung von Digitalisaten daher überhaupt sinnvoll?

I-6.4. Sollten bei einer qualitativ besseren Neuaufnahme diese Digitalisate an Stelle der alten Daten treten oder sollte hier ein neuer Datensatz angelegt und mit dem alten verlinkt werden?

[Meinungsfrage] Forschungsdaten müssen erhalten bleiben, um ihre Referenzierbarkeit und Nachvollziehbarkeit zu gewährleisten. Gilt dies auch für Digitalisate, die inhaltlich gleich sind, sich jedoch qualitativ unterscheiden?

I-6.5. Werden nicht mehr genutzte oder veraltete Digitalisate (z.B. veraltetes Dateiformat) aus dem System gelöscht?

- Wenn ja, unter welchen Bedingungen genau?

Persistenz ist eine wichtige Voraussetzung für die Nachnutzung von Forschungsdaten. Diese sollten daher erhalten bleiben, auch wenn sie veraltet sind.

I-7. Organisatorisches.

I-7.1. Liegt ein Datenmanagementplan oder ähnliches zum Umgang mit den Digitalisaten vor?

- Wenn ja, welche Aspekte des Datenmanagements werden beschrieben?
- Wenn ja, wer ist für die Umsetzung verantwortlich?

Für Forschungsdaten wird oftmals ein Datenmanagementplan gefordert, in dem der Umgang mit den Daten und die Verantwortlichen beschrieben werden. Wenn Digitalisate als Forschungsdaten zu verstehen sind, sollte solch ein Plan auch für diese vorliegen.

I-7.2. Wer ist Ihre Zielgruppe (Fachcommunity, interessierte Öffentlichkeit, andere Einrichtungen, ...)?

Die Frage soll klären, an wem sich die Institution bei ihren Angeboten am meisten orientiert.

I-7.3. Werden bekannte Nutzergruppen bei Entscheidungen (z.B. was als nächstes digitalisiert wird) mit einbezogen?

- Wenn ja, wie findet die Einbeziehung statt?

Die Digitalisate sollen, neben archivarisches Aspekten, vor allem die Zugänglichkeit zu den Informationen der Einrichtung erleichtern. Werden die Interessen der Nutzer dabei berücksichtigt?

I-7.4. Sollten Einrichtungen die Forschungsdaten, die aus ihren Digitalisaten hervorgehen (z.B. Korpora, Textauswertungen, ...), in Form eines Repositoriums sichern und verfügbar machen?

[Meinungsfrage] Ergebnisse, die aus Digitalisaten hervorgehen, könnten als weitere Forschungsdaten nachgenutzt werden. Wenn sie gemeinsam mit den Digitalisaten gespeichert und zugänglich gemacht werden, lassen sie sich vielleicht besser auffinden und werden so eher nachgenutzt.

I-7.5. Abschluss: Welche Aspekte sollten sich Ihrer Meinung nach bei der Digitalisierung in Gedächtniseinrichtungen verbessern?

[Meinungsfrage]

I-7.6. Abschluss: Welche Möglichkeiten (Digitalisierungstempo, Erschließungstiefe, Funktionen, ...) hätte Ihre Einrichtung bei unbegrenzten Mitteln?

[Meinungsfrage]

9.7.2. Interviewfragen der Nutzer mit Erläuterungen.

Ziel dieser Masterarbeit ist es zu klären, wie historische Quellen digitalisiert, verwaltet und bereit gestellt werden müssen, um den Anforderungen der Geschichtswissenschaft an ihre Forschungsdaten gerecht zu werden. Mit Hilfe von Interviews sollen dabei die Praxis und Möglichkeiten von Gedächtniseinrichtungen im Umgang mit Digitalisaten sowie die Anforderungen der Nutzer an diese Digitalisate untersucht werden.

Die Interviewfragen zu den Nutzern richten sich dabei nach einem schematischen Arbeitsablauf beim Suchen und Nutzen der Daten (Suche im System; Anforderungen an die Daten; Verfügbarkeit und Nachnutzung). Hinzu kommen Fragen zu den Wünschen, welche die Nutzer an die Angebote von Gedächtniseinrichtungen im Bezug auf Digitalisate haben.

Bei den Fragen hier handelt es sich um Faktenfragen, welche vor allem zur Einordnung und Vergleichbarkeit der befragten Nutzer erfasst werden.

Nf-1. System.

Nf-1.1. Ist Ihnen als Sucheinstieg die Suche oder das Browsing/Facettierung wichtiger?

Die Frage soll klären, ob Nutzer eher mit konkreten Suchen oder eher durch kontinuierliche Facettierung arbeiten.

Nf-2. Anforderungen.

Nf-2.1. Welche Metadaten eines Digitalisates sollten auf jeden Fall, welche idealerweise erfasst werden?

Die Frage soll klären, welche Metadaten die Nutzer für ihre Arbeiten brauchen und welche Metadaten die Gedächtniseinrichtungen daher auf jeden Fall erfassen sollten.

Nf-2.2. Welche Grundanforderungen stellen Sie an Bilddigitalisate (z.B. Farbigkeit, Vergrößerungsfunktion, Farbraum, Qualität, OCR, ...)?

Die Frage soll klären, welche Anforderungen die Nutzer an die Digitalisate stellen. Dies ist wichtig für einen Abgleich mit den Standards der Gedächtnisinstitutionen.

Nf-3. Verfügbarkeit und Nachnutzung.

Die Fragen zur Verfügbarkeit und Nachnutzung sind entweder detaillierter oder bauen aufeinander auf, so dass auf diese direkt im Interview eingegangen werden soll.

Nf-4. Wünsche.

Die Fragen zu den Wünschen der Nutzer sind umfangreicher und erfordern eine genauere Erklärung, so dass diese direkt im Interview beantwortet werden sollen.

Ziel dieser Masterarbeit ist es zu klären, wie historische Quellen digitalisiert, verwaltet und bereit gestellt werden müssen, um den Anforderungen der Geschichtswissenschaft an

ihre Forschungsdaten gerecht zu werden. Mit Hilfe von Interviews sollen dabei die Praxis und Möglichkeiten von Gedächtniseinrichtungen im Umgang mit Digitalisaten sowie die Anforderungen der Nutzer an diese Digitalisate untersucht werden.

Die Interviewfragen zu den Nutzern richten sich dabei nach einem schematischen Arbeitsablauf beim Suchen und Nutzen der Daten (Suche im System; Anforderungen an die Daten; Verfügbarkeit und Nachnutzung). Hinzu kommen Fragen zu den Wünschen, welche die Nutzer an die Angebote von Gedächtniseinrichtungen im Bezug auf Digitalisate haben.

N-1. System.

N-1.1. Nutzen Sie eher Portale (DDB, Europeana, ...) oder direkt die digitalen Angebote von Gedächtniseinrichtungen?

- Können Sie dies kurz begründen?

Die Frage soll klären, ob es für Gedächtniseinrichtungen sinnvoller ist, stärker ihre eigenen digitalen Angebote auszubauen oder eher auf Portale zu setzen.

Die Fragen N-1.1. bis N-1.2. dienen auch der Charakterisierung der Interviewpartner.

N-1.2. Welche Funktionen in digitalen Angeboten von Gedächtniseinrichtungen nutzen Sie am meisten bzw. sind am wichtigsten für Sie?

Es soll herausgefunden werden, ob bestimmte Funktionen wie Vergrößerung, Bearbeitungsmöglichkeiten, Downloadmöglichkeiten, etc. besonders genutzt werden.

N-2. Anforderungen.

N-2.1. Mit was für digitalen Daten arbeiten Sie hauptsächlich?

Zentrale Frage, die klären soll, mit welchen digitalen Daten die Nutzer überhaupt arbeiten. Daraus ergeben sich unterschiedliche Anforderungen an die Bereitstellung von Digitalisaten in Gedächtnisinstitutionen.

N-2.2. Favorisieren Sie bestimmten Metadatenformate/-schemata oder Normdaten für die Erschließung?

- Können Sie dies kurz begründen?

Durch die Frage soll in Erfahrung gebracht werden, ob sich die verwendeten Metadatenformate und -schemata von den hier untersuchten Gedächtnisinstitutionen und Nutzern decken.

N-2.3. Welche Informationen erwarten Sie von einer Dokumentation zu dem Digitalisat? (z.B. Provenienz, Informationen zum Digitalisierungsprozess, Informationen zum Digitalisierungsprojekt, ...)

Kontextinformationen sind ein wichtiger Aspekt bei der Nachnutzung von Forschungsdaten. Die Frage soll daher beleuchten, welche Informationen den Nutzern wichtig sind.

N-2.4. Welche Verarbeitungsstufe (Rohdaten, strukturiert, Volltext, ...) ist für Sie am sinnvollsten:

- bei Textdigitalisaten?
- bei Bilddigitalisaten?

Eine stärkere Aufbereitung der Daten durch die Gedächtniseinrichtungen bedeutet einen höheren Aufwand bei der Erstellung, aber auch mehr Verarbeitungsmöglichkeiten der Daten. Die Frage soll daher klären, bis zu welchem Grad die Nutzer die Daten aufbereitet brauchen.

N-2.5. Wie wichtig ist für Sie die Verlinkung von Objekten untereinander?

Eine Verlinkung von Objekten würde es den Nutzern erlauben, noch leichter Zusammenhänge zu erkennen und somit weitere Kontextinformationen liefern. Es könnten so auch zusammengehörige Objekte leichter gemeinsam ausgewertet werden.

N-2.6. Wie sollten Gedächtniseinrichtungen mit Dubletten umgehen: Beide Exemplare digitalisieren, um mögliche Unterschiede sichtbar zu machen, oder auf der Basis eines (oder mehrerer) Exemplare ein beispielhaftes Modell erstellen?

[Meinungsfrage] Hierdurch soll geklärt werden, ob für die Nutzer bei der Digitalisierung das Objekt oder der Inhalt im Vordergrund steht.

N-2.7. Durch die Digitalisierung gehen einige Informationen wie z.B. die Materialität der Quelle in der Nutzung verloren. Können umfangreiche Metadaten hier einen gewissen Ausgleich schaffen?

[Meinungsfrage] Informationen wie die Materialität gehen bei der Digitalisierung verloren, können in einigen Forschungsbereichen wie z.B. der Buchwissenschaft aber zentrale Untersuchungsgegenstände sein. Durch diese Frage soll die Meinung dazu herausgefunden werden, ob (und möglicherweise welche) Metadaten hier einen Ersatz leisten können.

N-3. Verfügbarkeit und Nachnutzung.

N-3.1. In wie fern beeinflusst die digitale Verfügbarkeit der Quelle Ihre Quellenauswahl?

Zum einen soll durch die Frage geklärt werden, ob Digitalisate als Quellen oder nur zum Kontrollieren analoger Quellen genutzt werden. Zum andern gibt die Frage einen Hinweis darauf, ob sich durch die Digitalisierung die Forschungspraxis dahingehend ändert, dass nicht mehr die analoge, sondern vor allem die digitale Verfügbarkeit für die Quellenauswahl und damit die untersuchte Forschungsfrage verantwortlich ist.

N-3.2. Vertrauen Sie Digitalisaten aus Gedächtniseinrichtungen als wissenschaftlichen Quellen?

- Wenn ja: Nur unter bestimmten Bedingungen?
- Wenn nein: Warum nicht? Was könnte eine Gedächtniseinrichtung hier verbessern?

Lange Zeit wurden Informationen aus dem Internet wegen fehlender Persistenz und Referenzierbarkeit als nicht geeignete Quellen betrachtet. Die Frage soll klären, wie die Nutzer in dieser Hinsicht den Digitalisaten aus Gedächtniseinrichtungen gegenüber stehen.

N-3.3. Wie sieht Ihr persönliches Forschungsdatenmanagement aus? Stützen Sie sich auf Angebote z.B. von virtuellen Forschungsumgebungen, nutzen Sie Cloud-Dienste oder arbeiten Sie lokal?

- Welche Dateien benötigen Sie hierfür?
- Gibt es Anforderungen aus der Fachcommunity oder von Förderern?

Die Frage soll die Arbeitsweise der Interviewten beleuchten. Daraus können Rückschlüsse für die Gedächtniseinrichtungen gezogen werden. (Z.B. sollte bei einer stark lokalen Arbeitsweise eine Downloadfunktion angeboten werden.)

N-3.4. Wie wichtig ist Ihnen eine Downloadfunktion für die angebotenen Dateien? Für einige Dateien besonders?

Die Downloadfunktion stellt eine Besonderheit dar, da sich die Daten somit auch lokal speichern und in anderen Tools als im digitalen Angebot der Gedächtniseinrichtung vorhanden verarbeiten lassen.

N-3.5. Benutzen Sie wenn möglich Harvesting, um große Datenmengen abzurufen?

Die Geschichtswissenschaft arbeitet überwiegend qualitativ, quantitative Auswertungen kommen eher in Spezialbereichen vor. Lohnt es sich daher für Gedächtniseinrichtungen, das Harvesten von Digitalisaten bzw. Metadaten anzubieten?

N-3.6. Wie beurteilen Sie Open Access für Digitalisate (sowohl bezüglich Zugänglichkeit wie Nachnutzung)?

[Meinungsfrage] Open Access gilt als wichtiger Aspekt im Forschungsdatenmanagement, da so Zugänglichkeit und Nachnutzung sichergestellt werden. Wie wichtig sind den Nutzern freie Zugänglichkeit sowie freie Nachnutzung von Digitalisaten aus Gedächtniseinrichtungen?

N-4. Wünsche.

N-4.1. Würden Sie als Nutzer gerne in Entscheidungsprozesse von Gedächtniseinrichtungen zur Digitalisierung einbezogen werden?

[Meinungsfrage] Die Digitalisate sollen, neben archivarischen Aspekten, vor allem die Zugänglichkeit zu den Informationen der Einrichtung erleichtern. Würden Nutzer dabei gerne an wichtigen Entscheidungsprozessen beteiligt sein?

N-4.2. Was sind für Sie die Forschungsdaten der Geschichtswissenschaft?

- Sind auch Metadaten Forschungsdaten?

[Meinungsfrage] Die Frage soll klären, was der Interviewte unter Forschungsdaten versteht, ob sich der Begriff mit der genannten Forschungsdatendefinition deckt oder ob

die Digitalisate vielleicht gar nicht als Forschungsdaten identifiziert werden. Die Frage ist Bewusst am Ende des Interviews gestellt, um vorher eine möglichst unvoreingenommene Einstellung zum Thema Digitalisate zu erhalten, sollten diese nicht mit Forschungsdaten assoziiert werden.

N-4.3. Wünschen Sie sich die Verwaltung von Forschungsdaten (Korpora, Textauswertung, ...) mit Bezug zu Digitalisaten durch die entsprechende Einrichtung?

- Wenn ja: Was spricht gegen ein disziplinäres Repositorium?
- Sehen Sie dies als Teil der Aufgaben von Gedächtniseinrichtungen?
- Würden Sie ihre Forschungsdaten selber in einem Repositorium ablegen, welches auch die Originale beinhaltet oder eher in fachspezifischen Repositorien?

[Meinungsfrage] Ergebnisse, die aus Digitalisaten hervorgehen, könnten als weitere Forschungsdaten nachgenutzt werden. Wenn sie gemeinsam mit den Digitalisaten gespeichert und zugänglich gemacht werden, lassen sie sich vielleicht besser auffinden und werden so eher nachgenutzt.

N-4.4. Abschluss: Können Sie kurz skizzieren, wie für Sie das ideale digitale Angebot einer Gedächtniseinrichtung aussehen würde?

[Meinungsfrage]

9.8. Gespiegelte Fragen.

<i>Gespiegelte Fragen</i>	<i>Institutionen</i>	<i>Nutzer</i>
Anforderungen an die Digitalisate	If-2.2	Nf-2.2
Welche Metadaten werden erfasst	If-3.1.	Nf-2.1.
Favoritisierte Metadatenformate	If-3.3	N-2.2
Verarbeitungsstufe/Volltext	If-3.5	N-2.4
Harvesting	If-5.1	N-3.5
Nutzung von Portalen	If-5.2	N-1.1
Dubletten	I-1.2.	N-2.6
Dokumentation schafft vertrauen	I-3.1	N-2.3
Originale als Ersatz/zusätzliche Metadaten	I-3.2	N-2.7
Open Access	I-4.1.	N-3.6.
Verlinkung	I-5.1	N-2.5
Beteiligung an Entscheidungen	I-7.3	N-4.1
Verwaltung der Forschungsdaten durch Institution	I-7.4	N-4.3
Funktionen	Webevaluation	N-1.2

9.9. Interviewtranskripte.

9.9.1. Interview Stiftung Stadtmuseum Berlin.

Stiftung Stadtmuseum Berlin - Sebastian Ruff/Sabine Weller

24.05.2019

If-2. Digitalisierung.

If-2.1. Werden bestimmte externe Standards zur Digitalisierung in Ihrer Einrichtung angewandt (z.B. die DFG-Praxisregeln zur Digitalisierung) oder folgt man hausinternen Richtlinien?

Die Standards sind an die DFG-Praxisregeln angelehnt bzw. bei Dienstleistern erfolgt die Digitalisierung nach DFG-Praxisregeln.

If-2.2. Welche Standards gelten konkret für die Erzeugung von Master-Bilddigitalisaten (z.B. DPI-Anzahl, Farbtiefe, Farbraum, Format, ...).

Als Handlungsanweisungen, auch an externe Dienstleister, ist angegeben: 300 dpi, TIFF-Format, RGB-Farbraum, 24 Bit Farbtiefe.

If-2.3. Wie viele Digitalisate bieten Sie insgesamt an?

In der Sammlung Online befinden sich 60610 Digitalisate. (Stand: 20.05.2019)

If-2.4. Wie viele Objekte werden pro Jahr ungefähr neu digitalisiert?

Ca. 18500 Objekte pro Jahr.

If-2.5. Befinden sich Born Digitals in Ihrer Sammlung?

Hier und da. In der Theatersammlung kommt es vor, dass Theaterfotografen Aufnahmen nur noch digital abgeben. Zudem ein bisschen Medienkunst (Inszenierungen), jedoch eher dokumentarische Quellen oder Objekte. Richtig digitale Kunst als neues Sammlungsgebiet ist noch nicht entstanden.

If-2.6. Findet die Digitalisierung im Haus oder durch Dienstleister statt?

- Wenn im Haus digitalisiert wird: Wie viele Mitarbeiter sind mit der Digitalisierung beschäftigt?

Sowohl als auch. Wenn man nur die Scankräfte und Datenerfasser/Inventur dazu zählt, dann konkret drei (Scanner, Fotograf, Datenerfasser). In den Prozess der Digitalisierung sind jedoch auch die Mitarbeiter der Sammlung, z.B. die Kuratoren, eingebunden, so dass man unter diesem Ansatz auf ca. 40 Mitarbeiter kommt.

Jedoch wird die Digitalisierung noch nicht als Aufgabenbereich aufgefasst, der über die normale Museumsdokumentation hinausgeht.

Massendigitalisierungen finden auch durch Dienstleister oder durch externe Fotografen statt.

If-2.7. Gibt es Maßnahmen, um Doppeldigitalisierungen von Objekten zu vermeiden?

- Finden Absprachen mit anderen Gedächtniseinrichtungen statt, um hier bereits digital vorliegende Werke nicht erneut zu digitalisieren?

In manchen Projekten ja, z.B. die Recherche, was es in anderen Sammlungen gibt. Besonders bei der Hausbibliothek ist dies ausgeprägt. Ansonsten ist unser Ziel jedoch, unsere eigene Sammlung digitalisiert zu haben. Wenn es ein Objekt bereits woanders digital gibt, ist dies also kein Ausschlusskriterium.

If-3. Dokumentation, Metadaten, Erschließungsstandards.

If-3.1. Welche Metadaten werden bei der Digitalisierung erfasst?

Als Pflichtfelder werden die Daten erfasst, die auch in der Sammlung Online angezeigt werden sollen: Inventarnummer, Sammlung, Titel, Künstler/Beteiligte, Zeit und Ort, Material und Technik, Maße, Schlagworte sowie idealerweise ein kurzer wissenschaftlicher Kommentar. Ein Datensatz lässt sich jedoch auch ohne diese Pflichtfelder (mit Ausnahme von Inventarnummer und Titel) speichern.

Ansonsten bietet Daphne viele weitere Möglichkeiten zur museologischen Erfassung:

- Sammlungsgebiete, in denen das Objekt vertreten ist.
- Kennzeichnungen des Objektes für die Onlinepräsentation. So lassen sich Objektgruppen auch über Sammlungsgrenzen hinweg kennzeichnen.
- Hinweise zu Markierungen oder Stempeln.
- Weitere beteiligte Personen und Orte.
- Bemerkungen sowie Literaturangaben zu dem Objekt.
- Daten zur Objektgeschichte (historische Sammlungsgebiete und Inventarnummern, Angaben zur Erwerbung und Provenienz).
- Daten zur Objektverwaltung (Standort, Leihstatus, Bearbeiter).
- Daten zur Restaurierung (Zustand, Bemerkungen zu durchgeführten Restaurierungen, Lagerungshinweise wie Beleuchtungsstärke, Luftfeuchte und Temperatur sowie Schadstoffbelastung).
- Registrardaten (Hinweise zu Transport und Verpackung, Versicherungen, Ausleihen)
- Ein Dokumentenarchiv mit Hinweisen zu verwandten oder in Verbindung stehenden Dokumenten und URLs.
- Ein Bildarchiv, in dem alle vom Objekt angefertigten Bilder sortiert sind.
- Weitere objektspezifische Daten, bei Münzen z.B. Nominalwert, Münzstätte, Beschreibungen zu Stempel und Rand.

If-3.2. Stehen alle erfassten Metadaten dem Nutzer frei zur Verfügung?

- Wenn nein: Welche nicht?

Stand heute nicht. Copyright-Angaben, Lizenzen und Schlagworte werden z.B. nicht, die wissenschaftliche Bearbeitung wird nur teilweise angezeigt.

Auf Anfrage wird natürlich Auskunft über alles erteilt, das System ist jedoch nicht so eingestellt, dass alles automatisch angezeigt wird.

Frei zur Verfügung stehen nur die Daten, die in der Sammlung Online angezeigt werden: Inventarnummer, Sammlung, Titel, Künstler/Beteiligte, Zeit und Ort, Material und Technik,

Maße sowie wenn vorhanden eine Anmerkung, die meist aus der wissenschaftlichen Bearbeitung stammt.

If-3.3. Nach welchen Metadatenschemata werden die Digitalisate beschrieben?

Die Daten werden in der Objektdatenbank Daphne erfasst. Diese ist nach dem SPECTRUM-Dokumentationsstandard zertifiziert. Bei den Bildern werden die EXIF-Daten ausgelesen. Von hier ist ein Export nach LIDO möglich.

If-3.4. Nach welchen Erschließungsstandards werden die Daten beschrieben? Kommen Normdaten zum Einsatz?

Die Objektdatenbank Daphne unterstützt die GND, GeoNames, AAT für Techniken.

If-3.5. Wird wenn möglich auch der Volltext erschlossen?

- Wenn ja, in welchem Format wird dieser gespeichert?

Ja, manchmal, wenn es für ein Forschungsprojekt benötigt wird. Der Volltext ist dann als Fließtext in die Datenstruktur der Daphne eingebunden. Manchmal liegt der Volltext auch noch als Word-Dokument vor.

If-4. Veröffentlichung.

If-4.1. Unter welchen Lizenzen sind die Digitalisate zugänglich?

- Welche Lizenz wird favorisiert? Warum?

Einige Objekte, welche für MuseumDigital verwendet wurden, sind mit Creative Commons Lizenzen ausgestattet. CC-0 und CC-BY kommen in der DDB vor. Der Großteil der Objekte unterliegt dem klassischen Copyright („Alle Rechte vorbehalten.“)

CC-Lizenzen wurden bisher nur für die Objekte vergeben, die in Portale geliefert werden. Inhouse ist dies noch ein offenes Thema.

Persönlich favorisieren wir die CC-Lizenzen je nach Einsatzzweck. Klar soll natürlich sein, in welcher Institution die Originale liegen (CC-BY). Möglich wären auch CC-0 oder die Public Domain. Die CC-Lizenzen schaffen Klarheit bei der Nachnutzung, transportieren nach außen, was gemeinfrei ist und gewährleisten bei der richtigen Lizenz trotzdem die Fotografennamensnennung. Es ist die sichtbare Umsetzung von Open Access. Da wir eine öffentliche Einrichtung sind, sollten wir diesem folgen.

If-5. Interoperabilität und Schnittstellen, Portale.

If-5.1. Können die Metadaten über maschinenlesbare Schnittstellen geharvestet werden?

- Wenn ja, welche Schnittstellen werden angeboten?

Nein, Schnittstellen und Harvesting gibt es nicht.

If-5.2. Werden die Digitalisate auch in andere Angebote oder Portale (z.B. die Deutsche Digitale Bibliothek/Europeana) eingespielt?

- Wenn ja, in welche Portale spielen Sie die Daten ein? Warum gerade in diese Portale?

- Wenn nein, warum nicht?

Digitalisate werden an die DDB gegeben, in der Europeana sind noch keine unserer Objekte, dies liegt jedoch an technischen Gründen der DDB. Zudem MuseumDigital, Kalliope, von der Bibliothek zudem im KOBV und der ZDB. Möglicherweise gibt es noch „Altlasten“ in Spezialdatenbanken, die von (ehemaligen) Sammlungskuratoren dort eingespielt, aber nicht zentral erfasst wurden.

Die DDB ist für die Förderung wichtig und bietet zudem die Chance, Objekte über Institutionsgrenzen hinaus sichtbar zu machen. MuseumDigital hängt an einem konkreten Projekt, zudem sind die einzelnen lokalen Instanzen besser geeignet als Unmengen an Metadaten und kuratierte Objekte lassen sich einfacher anbieten. Außerdem bietet das MuseumDigital gemeinsame Projekte mit anderen Institutionen und damit eine gute Vernetzung an, was gerade für den Berliner Fokus sehr wichtig ist.

Aktiv haben wir uns bisher nicht gegen Portale entschieden, die an uns herangetragen wurden.

If-7. Organisatorisches.

If-7.1. Welche Fördereinrichtungen (DFG, BMBF, EU, Land Berlin, ...) unterstützen Ihre Einrichtung?

Die DFG aktuell nicht, auch wenn es zukünftig wieder möglich ist. Viel über digiS durch das Land Berlin. Zudem die Koordinierungsstelle für den Erhalt schriftlichen Kulturguts (KEK), hier wurde zwar vor allem die Restaurierung bezahlt, welche jedoch als Vorschnitt zur Digitalisierung gesehen werden kann. EFRE-Mittel für den Aufbau der Website. Für einige Ausstellungen auch die Stiftung Deutsche Klassenlotterie.

If-7.2. Besteht die Möglichkeit, neue Funktionen in das System der digitalen Angebote zu integrieren?

- Wenn ja, wie oft geschieht dies?

Ja, die Möglichkeit besteht, über das Entwicklungsbudget. Neue Funktionen kommen im Schnitt einmal im Jahr. Das schließt aber auch komplett neue Versionen sowie neue Ansichtsfunktionen mit ein. Dies muss jedoch auch beim Entwickler beauftragt und beantragt werden.

If-7.3. Gibt es Erhebungen darüber, welche Nutzergruppen Ihre digitalen Angebote (am meisten) nutzen?

Für Sammlung Online verwenden wir Google Analytics. Es ist daher immer etwas schwierig, ob dahinter wirklich die Nutzergruppen stecken, die Google Analytics annimmt. Klar ist nur der IP-Bereich. Eine klassische Nutzerumfrage haben wir zur Sammlung Online noch nicht gemacht.

I-1. Auswahl.

I-1.1. Wie sieht der Auswahlprozess von Objekten für Digitalisierungsprojekte aus?

- *In wie fern sind Kuratoren/Sammlungsleiter und Konservatoren beteiligt?*
- *Welche Rolle spielen Schäden an den Objekten, z.B. besonders für oder gegen eine Digitalisierung?*
- *Fällt die Entscheidung vor allem durch Mittelgeber oder nach eigenen Bedürfnissen?*
- *In wie Weit werden die Nutzerbedürfnisse berücksichtigt bzw. die Nutzer eingebunden?*

Die Kuratoren sind dazu aufgerufen, Digitalisierungsprojekte einzureichen. Diese werden anschließend sondiert, ausgewählt und bei digiS eingereicht. Dies ist jedoch vor allem eine Leitungsentscheidung, was davon in welcher Reihenfolge für das Haus wichtig ist und was die Kollegen überhaupt schaffen können, da man oftmals dazu neigt, den Aufwand von Digitalisierungsprojekten auch für die betroffenen Kuratoren zu unterschätzen. So war für 2019 natürlich erstmal die Digitalisierung der Fontane-Manuskripte im Haus zentral. Im Vergleich zu anderen Häusern ist dies ein recht demokratischer Prozess, da Projekte vorgeschlagen werden können und dann erstmal eine Chance haben.

Schäden wird man versuchen, im Prozess der Digitalisierung gleich zu restaurieren. Beschädigte Objekte werden daher möglichst mit in die Digitalisierung übernommen, wenn sie in den inhaltlich Zusammenhang gehören. Es wurden jedoch auch Objekte aus Projekten rausgelassen, wenn dies aus kuratorischen und restauratorischen Gründen nicht zu verantworten war. Bisher gab es aber auch nur wenige Bestände, die zu zufallen drohten. Wenn dies der Fall war, so wurden diese Objekte natürlich auch zur Sicherungsdigitalisierung gegeben.

Die Entscheidung, was am Ende digitalisiert wird, fällt letzten Endes ausschließlich durch Mittelgeber, vorrangig digiS.

Nutzer werden aktuell nicht einbezogen, es wird so entschieden, wie man selber denkt, was die Nutzer brauchen. Man handelt also nach angenommenen Nutzerbedürfnissen. Ausnahmen bilden Aufträge durch Nutzer. Diese finanzieren dann die Digitalisierung, die Umsetzung findet aber weiter im Haus statt. Offene Befragungen, welche Bestände digitalisiert werden sollen, finden dagegen nicht statt.

I-1.2. Wie sieht der Umgang mit Dubletten aus: Werden beide Exemplare, z.B. eines Buches, digitalisiert, wird ein beispielhaftes Objekt aus beiden Dubletten erstellt oder wird sich für eines entschieden?

Dies ist projektbezogen. Bei einigen Projekten nimmt man das interessantere oder besser erhaltene Objekte, bei anderen alle.

Wir folgen hier pragmatischen Fragen, es gibt keine Regel, sondern wir besprechen dies bei jedem Projekt neu. Dies ist natürlich auch eine Geldfrage. Bei Projekten mit vielen Dubletten wird man wahrscheinlich nicht jedes Objekt digitalisieren.

Zudem gibt es museologische Dubletten mit gleicher Inventarnummer, die jedoch keine eigentlichen Dubletten sind, z.B. Drucke mit unterschiedlichen Papierarten. Hier ist eine Digitalisierung beider Objekte ein echter Mehrwert, trotz der selben Inventarnummer sind es hier zwei Objekte, dies könnte auch von Forschungsinteresse sein.

Grundsätzlich tendiert man aber dazu, ein Objekt auszuwählen und auf die Dubletten hinzuweisen.

I-2. Digitalisierung.

I-2.1. Erfolgt die Digitalisierung immer vom Original oder wenn vorhanden auch vom Film/Microfiche?

Die Digitalisierung von Ektachromen wurde mal versucht, die Ergebnisse hier waren aber zu schlecht.

Grundsätzlich erfolgt die Digitalisierung immer vom Original. Bei Fotos wird vom Abzug oder manchmal dem Negativ digitalisiert. Natürlich würde die Kuratorin der Fotosammlung sagen, vom Abzug sollte man nicht digitalisieren, da dieser bereits ein Derivat darstellt und hier bereits Verlust drin ist, man also einen Abzug vom Abzug macht, dennoch wird es bei Fotos manchmal gemacht, wenn wir das Negativ nicht haben.

Sonst entscheiden wir uns in der Regel wegen der Qualität für das Original, besonders auch, da unsere Microfiches relativ alt sind.

Generell immer vom Original kann man daher nicht sagen, aber immer so original wie möglich.

I-2.2. Als Qualitätsstandard werden oft die DFG-Praxisregeln zur Digitalisierung genannt. Wie relevant sind diese für die Digitalisierung in ihrer Einrichtung?

Die Verwendung wird bei Dienstleistern abgefragt, intern sollten wir unseren Workflow aber noch mehr auf die Standards ausrichten.

I-2.3. Welche Maßnahmen zur Qualitätssicherung werden getroffen?

Die optische Qualität und Sichtkontrolle findet in der Fotothek statt. Was wir nicht machen ist die Qualitätskontrolle der Digitalisate mit kalibrierten Monitoren am Ort wo die Originale liegen. Wenn die Digitalisierung bei einem Dienstleister geschieht, so kontrolliert dieser, ob das Objekt vollständig ist oder es krasse Farbstiche gibt. Die Farbechtheit wird mittels ColorChecker im Workflow kontrolliert.

In der Fotothek findet eine Überprüfung zum einen beim Einspielen der Daten statt, hier wird auf Unschärfe, Farbfehler etc. geachtet. Bei fehlerhaften oder nicht kompatiblen Farbprofilen wird zudem die Daphne diese beim Einspielen zurückweisen. Zum andern findet eine laufende Kontrolle statt. Wenn Objekte durch Anfragen benötigt werden, wird zuerst geprüft, ob die vorhandenen Digitalisate noch den aktuellen Anforderungen entsprechen. Diese sind im Moment 300 dpi im TIFF-Format, Verwendung des RGB-Farbraums sowie eine Farbtiefe von 24 Bit.

I-2.4. Werden Daten im Nachhinein noch verändert?

- Wenn ja, findet eine Erfassung der Veränderungshistorie statt?

Grundsätzlich nein. Es kann jedoch sein, dass einige Bilder noch im Digitalisierungsprozess nachbearbeitet werden. Ziel des Digitalisierungsworkflows ist es jedoch, bereits in der Digitalisierung so viel Qualitätssicherung und Farbechtheit zu haben, dass das Ergebnis ohne Kontrolle sicher dem Original so ähnlich wie möglich sein

kann. Die Bilder werden so an die Nutzer gegeben, welche die Digitalisate dann für ihre Ansprüche anpassen können.

Sollte doch verändert werden, so dokumentieren wir dies jedoch nicht. Auch neue Bearbeitungsstufen werden nicht als neue Fotos abgespeichert.

I-3. Dokumentation, Metadaten, Erschließungsstandards.

I-3.1. Welche Informationen zum Kontext der Digitalisierung (Sammlungsbereich, Provenienz, Digitalisierungsprojekt, ...) werden dokumentiert und verfügbar gemacht?

Wir haben ein Mindestset an Daten, welches ein Sammlungsobjekt in der Vorbereitung der Digitalisierung haben muss. Dies betrifft eher museologische Grunddaten, welche auch in der Sammlung Online auftauchen: Inventarnummer, Sammlung, Titel, Künstler/Beteiligte, Zeit und Ort, Material und Technik, Maße sowie wenn vorhanden eine einfache wissenschaftlichen Beschreibung.

Manche Projekte sind Forschungsprojekte, die Digitalisierung ist hier Anlass, die Objekte zu erschließen und in einen Kontext zu setzen. Gerade wenn man die Digitalisate nicht verstehen kann, ohne den historischen Kontext zu kennen. (Z.B. bei historischen Flugblättern, wo man wissen muss, welche Parteien/Personen angegriffen werden, um die Flugblätter richtig zu verstehen.)

Ansonsten machen wir weniger Tiefenerschließung sondern vor allem Massendigitalisierungsprojekte. Die Metadaten werden von den Sammlungskuratoren hier nur fließbandartig aufgenommen oder die Erfassung erfolgt sogar durch Dienstleister. Was wir in den letzten Jahren nie gemacht haben ist die systematische Transkription und die Erfassung von Strukturdaten von mehrseitigen Objekten, auch wenn wir hier aktuell gerade dran sind, wenigstens Paginierung und Kapitel in den Metadaten zu erfassen.

Provenienzen werden nicht anlässlich der Digitalisierung erforscht, dies geschieht getrennt in der Provenienzforschung.

Die Erfassung der zugehörigen Digitalisierungsprojekte erfolgt über die Möglichkeit der Kennzeichnung bei der Darstellung.

I-3.2. Durch die Digitalisierung gehen einige Informationen wie z.B. die Materialität der Quelle in der Nutzung verloren. Können umfangreiche Metadaten hier einen gewissen Ausgleich schaffen?

Klar könnten Metadaten mehr dazu beitragen, was das echte Objekt ausmacht. Die Frage ist, ob man das will. Ich würde nie den Anspruch erheben, so viele Metadaten zu erfassen, dass man sagen kann, man braucht das Original nicht mehr, da die Metadaten alles über das Original sagen. Sicher gibt es Forschungszusammenhänge, wo man noch mehr zu den Objekten erfassen könnte, um sie irgendwie begreifbar zu machen. Aber es ist nicht der Anspruch und wird nie möglich sein, eine vollständige digitale Repräsentanz einschließlich Haptik zu haben. Dies ist auch nicht der Anspruch von Digitalisierung im Museum, das Original zu ersetzen, sondern das Ziel ist es, sichtbar zu machen, was das Museum besitzt.

I-4. Veröffentlichung.

I-4.1. Werden alle digitalisierten Objekte (soweit rechtlich möglich) auch in Ihren digitalen Angeboten zur Verfügung gestellt? Sind hier alle angebotenen Digitalisate frei zugänglich?

Frei zugänglich ist interpretationswürdig. Wahrscheinlich nein und nein, wir haben eine Ratio von 1:5, 300.000 Datensätze zu ca. 60.000 Objekten in der Sammlung Online. Es gibt bei uns keinen Automatismus der besagt digitalisiert ist gleich veröffentlicht. Das liegt zum einen an der Qualitätsprüfung, dass wir sagen, Metadaten und Fotos sollen einen gewissen Standard erfüllen, und es liegt zum andern daran, dass wir den Veröffentlichungsmechanismus noch nicht demokratisiert haben. Es liegt noch nicht vollständig in der Hand der Kuratoren, welche ihrer Objekte veröffentlicht werden. Zudem sollte auch keine Veröffentlichung ohne Foto erfolgen.

Frei zugänglich ist zudem etwas anderes, als wir es gerade anbieten. Die Digitalisate sind nicht in hoher Auflösung herunterzuladen, die Metadaten sind nicht über eine API oder ein Repositorium zugänglich. Wir sind sichtbar, die Daten sind aber nicht frei zugänglich.

I-4.2. Werden die Digitalisate in der höchstmöglichen Qualität (Master) in der digitalen Sammlung Ihrer Einrichtung angeboten?

- Wenn nein, warum nicht, worin unterscheiden sich die angebotenen Digitalisate vom Master?

- Wenn nein, was ist nötig, um die Master-Dateien zu erhalten?

Nein. Die TIFFs sind über den Viewer der Sammlung Online sichtbar, jedoch kann man sie nicht runterladen. Zudem gibt es viele Datensätze, bei denen die TIFFs nicht angeboten werden. Es wird also nicht generell die höchste Qualität die vorliegt auch angeboten. Dies liegt zum einen daran, dass von einigen Objekten nur JPEGs vorhanden sind, die sich dann in geringerer Auflösung und Kantenlänge unterscheiden. Dies liegt auch daran, dass wir keine Ressourcen haben, für alle Objekte TIFFs zu erstellen. Viele Bilder stammen von den Mitarbeitern der Sammlungen, welche die Objekte mit normalen Kameras fotografieren.

Wenn TIFFs zum andern vorliegen, aber nicht veröffentlicht werden, liegt dies teils an Speicherplatzfragen, teils daran, dass die damit verbundenen Möglichkeiten von Zoom etc. auch sinnvoll für ein weiteres Verständnis des Objektes sein müssen. Hier findet also wieder ein gewisser Auswahlprozess statt.

Ein weiterer Grund ist die Vermarktungsmöglichkeit, welche die Bilder bieten. Dieser stellt zudem eine Kontrollmöglichkeit dar, wo unsere Bilder verwendet werden. Hinzu kommen Lizenzfragen, einige Bilder dürfen wir nicht oder nur mit einer geringen Auflösung frei zugänglich bzw. sichtbar machen.

Durch direkte Kontaktaufnahme erhalten die Nutzer die Master-Dateien, teilweise gegen Gebühren.

I-4.3. Welche Rolle spielen Zertifikate für Forschungsdatenrepositorien (z.B. Data Seal of Approval oder das DINI-Zertifikat) bei Ihren digitalen Angeboten?

Gar keine. Mit diesem Thema haben wir uns noch nicht beschäftigt, bisher war es aber auch noch nicht nötig.

I-5. Interoperabilität und Schnittstellen, Portale.

I-5.1. Findet eine Verlinkung zwischen Objekten (intern wie extern) statt?

In der Daphne ja, über zwei Wege. Zum einen durch die Verwendung der selben normierten Begriffe in bestimmten Feldern, so dass über eine Suche zumindest zusammenhängende Objekte mit den gleichen Begriffen gefunden werden. Zum andern gibt es die Funktion der zugehörigen Objekte, eine direkte Verlinkung in der Datenbank, wo man Angeben kann „ist gleichrangig“ oder „ist Teil von“.

Online findet die Verlinkung über die Visualisierung in den selben Datenfeldern statt. In der neuen Datenbank-Version gibt es auch eine einfache Vorschlagsfunktion, die sich auf die selben Metadaten (Sammlung, Schlagworte, ...) bezieht. Dies ist jedoch nicht visualisiert.

Wir verlinken sehr selten aktiv, vieles passiert einfach durch normierte Begriffe.

I-5.2. Werden die Digitalisate von Ihrer Seite auch auf weniger wissenschaftlichen Plattformen wie Wikipedia angeboten?

- kurze Begründung

Ja, wir haben einigen Content in der Wikipedia, der jedoch als Teil von Hackathons (Coding da Vinci) dorthin gelangt ist. Zudem haben wir einige Artikel geschrieben, diese jedoch kaum mit Digitalisaten bespielt.

Wir würden uns aber natürlich nicht verwehren, wenn jemand dafür anfragt. Wir würden gerne mal wieder mit Wikipedia oder Wikimedia ein Commons Projekt machen, mit Freiwilligen arbeiten und Digitalisate erstellen. Es finden immer wieder mal Treffen statt, wie Kultureinrichtungen und Wikimedia zusammenarbeiten könnten.

Aktuell sind nur ein paar duzend Objekte in der Wikipedia eingespielt, aber es gibt durchaus noch mehr Objekte, die dafür geeignete Lizenzen hätten.

In Social Media, z.B. Pinterest, werden Digitalisate für redaktionelle Zwecke, also Museumskommunikation, verwendet. Wenn Bilder verwendet werden, dann nicht um das Bild anzubieten, sondern um auf Veranstaltungen/Ausstellungen hinzuweisen. Es ist nicht so, dass man hier Daten zur Verfügung stellt.

I-6. Speicherung, Pflege und Archivierung.

I-6.1. Gibt es spezielle Strategien (z.B. regelmäßige Migration, Konvertieren, ...) zur Langzeitarchivierung der Daten?

Im Haus haben wir keine Prozesse zur LZA. Wir arbeiten vor allem mit Backup-Strategien um sicherzustellen, dass keine Speichermedien kaputt gehen. Ansonsten speichern wir redundant und auf Band. Die Sicherungskopien finden dabei wöchentlich statt.

Langzeitarchivierung wird beim Zuse Institut eingekauft, aktuell auch ausschließlich für Digitalisierungsprojekte aus Mitteln des Zuse Instituts. Probleme der LZA haben wir also komplett an Zuse ausgelagert.

I-6.2. Wie sieht die Backup-Strategie für die Digitalisate aus?

- In welchen Abständen werden Sicherungskopien durchgeführt?

- Werden redundante Sicherungskopien geführt?
- Werden sie in anderen Formaten als im laufenden System gesichert?

Wir sichern manchmal die Masterdateien der Digitalisierung nochmals separat, z.B. in einer höheren Auflösung oder als RAW-Format. Dies sind dann aber ausschließlich Datenformate zur Speicherung, geschieht aber nicht automatisch. Hauptsächlich speichern wir die Daten, mit denen wir auch arbeiten.

I-6.3. Lohnt sich eine langfristige Archivierung von Digitalisaten (im Bezug auf Aufwand, Finanzen, Verwaltung, ...) oder sollten diese bei Bedarf (beschädigte Daten, veraltete Formate, ...) einfach neu vom Original aufgenommen werden?

Natürlich sollte immer erstmal in eine ordentliche LZA investiert werden, zumal es für uns als Stadtmuseum bei Zuse nichts kostet, da das Land Berlin dies zahlt.

Zudem spiegelt der aktuelle Zustand wahrscheinlich auch einen besseren Zustand des Originals wider, wer weiß, wie das Original in 30 Jahren aussieht. Ein Digitalisat ist ja immer auch irgendwie ein Zustandsfoto.

Durch Standards bei der LZA sollte verhindert werden, dass neu vom Original aufgenommen werden muss. Was man natürlich nicht weiß ist, ob sich irgendwann mal neue Industriestandards für die Digitalisierung von Objekten durchsetzen, so dass wir neu digitalisieren müssen. Früher arbeitete man mit schwarz-weiß Microfiches, heute will man farbige Microfiches.

Gerade auch wenn man den Berg an Objekten betrachtet, der noch nicht digitalisiert ist, würde man erstmal davon ausgehen, ein Objekt einmal zu digitalisieren und dabei nicht im Hinterkopf zu haben, dieses bald wiederholen zu müssen. Das ist auch eine Frage der Effektivität.

I-6.4. Sollten bei einer qualitativ besseren Neuaufnahme diese Digitalisate an Stelle der alten Daten treten oder sollte hier ein neuer Datensatz angelegt und mit dem alten verlinkt werden?

Im Idealfall sollten beide Digitalisate erhalten werden, auch um sie als Zustandsfotos zu haben. Dazu muss man sich aber den benötigten Speicher leisten können.

I-6.5. Werden nicht mehr genutzte oder veraltete Digitalisate (z.B. veraltetes Dateiformat) aus dem System gelöscht?

- Wenn ja, unter welchen Bedingungen genau?

Dieser Fall ist bisher bei uns noch nicht eingetreten, dafür digitalisieren wir noch nicht lange genug. Unsere Formate (JPEG, TIFF, MP3, WAV, ...) sind alle noch lesbar.

Aber wir löschen z.B. keine JPEGs, nur weil nun ein TIFF vorliegt.

I-7. Organisatorisches.

I-7.1. Liegt ein Datenmanagementplan oder ähnliches zum Umgang mit den Digitalisaten vor?

- Wenn ja, welche Aspekte des Datenmanagements werden beschrieben?
- Wenn ja, wer ist für die Umsetzung verantwortlich?

Nein, ein formalisierter Ablaufplan liegt noch nicht vor, muss aber kommen. Dies ist eine Gemeinschaftsaufgabe aller, die daran beteiligt sind, also Daten erstellen, speichern, sichern und bereitstellen.

I-7.2. Wer ist Ihre Zielgruppe (Fachcommunity, interessierte Öffentlichkeit, andere Einrichtungen, ...)?

Auf jeden Fall die Forscher in speziellen Sammlungen, also die jeweilige Fachcommunity. Zudem andere Kolleginnen und Kollegen aus anderen Museen, Studierende und Dozenten, die interessierte Öffentlichkeit. Eine große Gruppe sind zudem private Laienforscher sowie kommerzielle Nutzer.

I-7.3. Werden bekannte Nutzergruppen bei Entscheidungen (z.B. was als nächstes digitalisiert wird) mit einbezogen?

- Wenn ja, wie findet die Einbeziehung statt?

Nein, aktuell nicht, außer sie beauftragen selber eine Digitalisierung und geben dazu Geld.

I-7.4. Sollten Einrichtungen die Forschungsdaten, die aus ihren Digitalisaten hervorgehen (z.B. Korpora, Textauswertungen, ...), in Form eines Repositoriums sichern und verfügbar machen?

Wenn uns jemand Metadaten oder Forschungsdaten anbietet, die wir noch nicht haben, sollten wir dies auf jeden Fall in unser Erschließungssystem aufnehmen.

Wir sollten jedoch nicht die Rolle eines weiteren Portal für die Fachcommunity übernehmen, sondern eher fragen, ob bereits Daten dazu bestehen und wir unser Digitalisat, was jemand angereichert hat, mit diesen Daten zusammenführen können. Die Gefahr besteht, dass man denkt, man müsse das nächste Fachportal sein. Lieber sollte man die Infrastruktur nutzen, die bereits von Forschern besteht und genutzt wird. Bis man selber so weit ist, anerkannte Quelle zu sein, dauert es viel zu lange und dann wird man doch nicht gefunden.

Gerne übernehmen wir einen Link zu den Forschungsdaten unserer Digitalisate und würden vielleicht auch Daten wie die Volltexte einspielen, sonst aber die Daten dort zusammenführen, wo die entsprechenden Forscher auch nachschauen.

Zwischenfrage: Sollte man auch aktiv auf eine Verlinkung der Daten hinweisen?

Klar, damit die Quelle sichtbar bleibt sollte man hier auch aktiv werden.

I-7.5. Abschluss: Welche Aspekte sollten sich Ihrer Meinung nach bei der Digitalisierung in Gedächtniseinrichtungen verbessern?

1. Verfügbarmachung und der Standard von Open Access sollten eingeführt und danach gehandelt werden.
2. Standardworkflows und Qualitätsstandards auch bei Digitalisierungen im Haus sollten eingeführt werden. Keine Sonderlösungen, die nur für einzelne Institutionen gelten, sowohl was Ablaufworkflow, Qualität, Verfügbarmachung und Datenformate betrifft.
3. Nicht für alles das neue Portal sein wollen, sondern schauen, wo man hinverlinken und wie man andere Portale nutzen und die eigenen Objekte zur Verfügung stellen kann.

4. Digitalisierungsprojekte noch mehr auf den Nutzwert hin abklopfen. Digitalisieren wir nur für uns selbst oder auch weil wir wissen, dass es hier eine Nachfrage gibt.

5. Eigene Ressourcen aufbauen. Wenn man Qualität und Quantität haben möchte, bedarf es einer Investition. Man sollte sich darüber bewusst sein, dass man ein Projekt nicht umsonst kriegt und sollte sich nicht darauf verlassen, dass Drittmittel die Sammlungsdigitalisierung sichern. Die Kernaufgabe Digitalisierung darf nicht davon abhängig sein, dass der nächste Förderantrag erfolgreich ist, sondern muss bis in die Geschäftsleitung als Gesamtauftrag der Institution verankert sein. Jeder muss verstehen, welchen Teil er zu dieser Strategie beiträgt und diesen Teil auch leisten. Das bedarf auch der nötigen Stellen und Technik in der Institution.

I-7.6. Abschluss: Welche Möglichkeiten (Digitalisierungstempo, Erschließungstiefe, Funktionen, ...) hätte Ihre Einrichtung bei unbegrenzten Mitteln?

Bei unbegrenzten Mitteln hätten wir die Chance, unseren gesamten Sammlungsbestand online zu zeigen und zur Verfügung zu stellen. Wir hätten die Möglichkeit, unseren schrumpfenden Personalkörper dadurch auszugleichen, dass wir mehr Mitarbeiter beteiligen an Erschließung und Digitalisierung. Wir hätten die Möglichkeit, viel sichtbarer zu werden, indem wir mehr zeigen und aktiver zu werden in Portalen und für Nutzergruppen, die wir uns heute noch gar nicht vorstellen.

Und man könnte ganz neue Aspekte der Präsentation für ein Museum etablieren, in denen man mit Digitalisaten und nicht nur mit Originalen arbeiten und mit den Objekten im digitalen Raum aktiver wird.

Das Museum hätte die Möglichkeit, relevant zu sein für die Nutzer des 21. Jahrhunderts, wirklich einen Mehrwert und Content zu bieten, der die hohen Ausgaben der Digitalisierung begründet. Eine verlässliche Quelle für Wissen und Daten. Daraus würde sich auch die Möglichkeit ergeben, im Netz aktiv Forschung betreiben zu können und inhaltliche wertvolle und nachhaltige Akzente in der digitalen Welt zu setzen.

Bei unbegrenzten Mitteln wären wir in zehn Jahren mit der Digitalisierung fertig. Selbst mit 1000 Mitarbeitern würde man so lange brauchen, da die Infrastruktur bleibt. Aber danach wäre man fertig und könnte sich wieder auf andere Museumsaufgaben konzentrieren und sich neu aufstellen. Wenn man alles online hat, dann kann man auch die Ressourcen der Sammlung wieder mehr in Forschung stecken und wäre somit wieder mehr ein forschendes Museum.

Die Erschließungstiefe sollte dabei gar nicht so sehr an erster Stelle stehen. Hier könnte man mehr die Community mit einbeziehen, denn wenn man sichtbar ist, dann können auch andere mit den Daten forschen. Die Ausschließlichkeit, die Erschließung im Museum zu belassen, müsste dann gar nicht mehr bestehen bleiben.

If-2. Digitalisierung.

If-2.1. Werden bestimmte externe Standards zur Digitalisierung in Ihrer Einrichtung angewandt (z.B. die DFG-Praxisregeln zur Digitalisierung) oder folgt man hausinternen Richtlinien?

Sowohl als auch. Die DFG-Standards sind auf jeden Fall die maßgebliche Leitlinie für die Digitalisierung, werden seit Anfang an angewandt und definieren für uns auch den Mindeststandard. Diese Funktion ist ganz wichtig. Wir gehen also auf keinen Fall, selbst wenn ein Projekt nicht DFG-gefördert ist, unter diese Anforderungen. Sie garantieren uns einen Standard, dass wir mit anderen Einrichtungen vergleichbar sind, einen Mindeststandard haben und eine gewissen Interoperabilität der Daten vorliegt. Zudem bilden sie eine Sicherheit, die erzeugten Digitalisate soweit machbar über lange Zeit zu präsentieren und mindestens über den DFG-Viewer sichtbar zu halten, wodurch die Verbreitung und Sichtbarkeit garantiert ist.

Wir gehen aber auch über die DFG-Richtlinien hinaus. Z.B. haben wir bei vielen Projekten für jedes Projekt eigene Projektrichtlinien beim Scannen, in denen wir bestimmte Feinheiten festlegen. Das kann, je nachdem, was gescannt wird, eine Erhöhung der Auflösung sein, aber auch Anforderungen wie bestimmte Bestände mit oder ohne Rand scannen oder ob man den Buchschnitt mitscannt oder nicht. Da sind die DFG-Richtlinien ja erstmal offen, daher wird sowas in den Projektrichtlinien festgelegt, je nachdem, was sich für die Objekte anbietet.

Darüber hinaus versuchen wir mehr und mehr die Metamorfoze-Guidlines bzw. den ISO 19264-1-Standard zu berücksichtigen, also ein gewisses technisches Mindestmaß an die Abbildungsqualität der Scanner sicherzustellen. Das ist nicht in den DFG-Standards festgelegt, aber wichtig, damit man die Originalgetreue des Werkes möglichst gut erfassen und perspektivisch nachweisen kann, diese zu garantieren.

Wenn wir nicht nur die Images betrachten, dann haben wir auch selbstgewählte Standards für die Strukturdatenerfassung. Diese werden projektspezifisch ausgewählt von gar nichts bis hin zu feinstens strukturierter Erfassung, wo auch eine Auswahl an projektspezifischen Benennungen zur Verfügung steht. Das ist aber der einzige Bereich der Erfassung, in dem Standards für die Digitalisierung festgelegt sind. Für den Nachweis gelten bei uns die selben Standards, die wir für die Katalogisierung anwenden.

If-2.2. Welche Standards gelten konkret für die Erzeugung von Master-Bilddigitalisaten (z.B. DPI-Anzahl, Farbtiefe, Farbraum, Format, ...).

Minimalanforderungen sind mindestens 300 dpi, 24 Bit Farbtiefe, Farbraum RGB, Masterformat ist unkomprimiertes TIFF.

Von da aus kann man natürlich in bestimmten Fällen nach oben abweichen. Bspw. hat unsere Mittelformatkamera in der Fotostelle 16 Bit Farbtiefe, die dpi-Anzahl wechselt je nach Projekt, mal 300, mal 600, mal 400 dpi. Es wird immer eine TIFF-Datei erzeugt, im Rahmen von DoD können jedoch auch JPEG oder PDF ausgegeben werden.

Wenn von Microfilmen digitalisiert wird, z.B. in der Zeitungsdigitalisierung, ist der Farbraum eine Graustufenaufnahme oder Bitmap, wenn die Vorlage nichts anderes erlaubt bzw. nichts anderes hergibt.

If-2.3. Wie viele Digitalisate bieten Sie insgesamt an?

Aktuell (Stand 19.06.2019) befinden sich 154.159 Werke in den Digitalisierten Sammlungen. Wir zählen dabei die Goobi-Vorgänge, wobei ein Vorgang sowohl ein umfassendes Werk mit vierstelliger Seitenzahl als auch ein einseitig beschriebener Brief sein kann. Wenn man die Images zählt haben wir ca. 20 Mio. Dateien.

If-2.4. Wie viele Objekte werden pro Jahr ungefähr neu digitalisiert?

Im letzten Jahr hatten wir 1,4 Mio. Images an den Scannern sowie 600.000 Images über den Microrollfilm-Scanner, haben also die Zahl von 2 Mio. erreicht. Dies liegt aber an dem Zeitschriften-, also Rollfilm-Projekt, in der Regel sind wir zwischen 1,0 und 1,5 Mio. Images, die an den Scannern produziert werden. An Vorgängen wechselt das immer sehr stark, meistens aber im Bereich zwischen 10.000 und 15.000 Vorgängen.

If-2.5. Befinden sich Born Digitals in Ihrer Sammlung?

In der digitalen Sammlung nicht, in diesen wird nur dargestellt, was das Digi-Zentrum wirklich herstellt. Ansonsten haben wir solchen Material natürlich im Haus, aber nicht in dieser Infrastruktur, sondern diese werden in eigenen Repositorien gespeichert. In Zukunft ist geplant, diese zwar auch über eine OAI-Schnittstelle anzubieten, aber nicht in die digitale Sammlung zu überführen. Das heißt, man findet Born Digitals im Katalog, aber nicht unter den digitalen Angeboten unter digital.staatsbibliothek-berlin.de.

If-2.6. Findet die Digitalisierung im Haus oder durch Dienstleister statt?

- Wenn im Haus digitalisiert wird: Wie viele Mitarbeiter sind mit der Digitalisierung beschäftigt?

Wir digitalisieren wenn irgendwie möglich immer im Haus. Früher, als es das Digitalisierungszentrum noch nicht gab, wurden auch externe Dienstleister beauftragt, davon geht man aus Bestandsschönungsgründen aber immer weiter zurück, zumal sich auch die Tendenz zeigt, dass die Bestände immer spezieller werden. Wir kommen immer mehr und mehr zu handschriftlichen Materialien oder Nachlässen in der Digitalisierung. Dies ist prinzipiell kein Hindernis, aber eine Vorsichtsmaßnahme, dass wir die Digitalisierung lieber wenn möglich im Haus durchführen. Wir haben aber auch Fälle, bei denen wir nicht die technische Ausstattung haben. Hier wird dann eine Ausschreibung vorbereitet und ein Dienstleister beauftragt. Das stellt aber nur einen minimalen Teil der Digitalisierung dar.

Im Digitalisierungszentrum (Abteilung II F Referat 3) direkt arbeiten immer zwischen 20 und 25 Personen. Der gesamte Digitalisierungsprozess bildet sich aber über das gesamte Haus ab, wenn man auch die Kollegen in den Abteilungen mitzählt, die die Bestände vorbereiten, erschließen, Strukturdaten vergeben, in der Datenverwaltung arbeiten etc. dann kommt man auf über 100. Natürlich sind dabei nicht alle in Vollzeit mit der Digitalisierung beschäftigt, aber sie sind beteiligt.

If-2.7. Gibt es Maßnahmen, um Doppeldigitalisierungen von Objekten zu vermeiden?

- Finden Absprachen mit anderen Gedächtniseinrichtungen statt, um hier bereits digital vorliegende Werke nicht erneut zu digitalisieren?

Ja. Zum einen beim Fall DoD. Hier gibt es immer zuerst eine Prüfung, ob das Digitalisat bereits vorhanden ist oder ob es eventuell in anderen, naheliegenden Bibliotheken bereits vorhanden ist. In diesem Fall würden wir im Katalog dann darauf verweisen, damit sich die Frage im Optimalfall nur einmal stellt. Dabei haben wir aber eine strikte Auswahl, jedes Digitalisat ist auch nicht gut genug. Die Kollegen in der Benutzungsabteilung haben vor einigen Jahren mal eine kleine Rangliste erstellt, welche Anbieter gute, zuverlässige, vollständige Digitalisate bietet. Für den Fall, dass ein Digitalisat uns nicht gut genug erscheint, würden wir dann wahrscheinlich doch neu digitalisieren.

Zum andern kennzeichnen wir unseren Bestand, der digitalisiert ist, so dass man punktuell feststellen könnte, dass ein Werk bereits digital vorliegt, da dann bereits ein Digitalisierungsvermerk im Band enthalten ist.

In den Großprojekten wie VD 16, 17, 18 gab es auch Phasen, in denen man bestimmte Bestandsgruppen bestimmten Bibliotheken zugeteilt hat. Man versucht eben, Doppeldigitalisierung zu vermeiden und einen effizienten Mitteleinsatz zu garantieren.

Auf der anderen Seite merkt man aber auch, dass dies nicht immer möglich ist. Im Digitalisierungsprojekt zum Ersten Weltkrieg bspw. waren auf Grund des bevorstehenden Jahrestages zeitgleich zahlreiche Digitalisierungsprojekte aktiv, die unabhängig voneinander entstanden sind und erst während der Laufzeit aufeinander aufmerksam wurde. So ergab sich z.B., dass sich die Digitalisierungsprojekte der DNB und der Staatsbibliothek teilweise in den Beständen überschneiden. Ein nachträglicher Abgleich hätte hier dann jedoch einen zu hohen Aufwand bedeutet. Der selbe Effekt entstand, als die Massendigitalisierung richtig Fahrt aufnahm. Anfangs hatte die einzelnen Institutionen nur auf ihren Bestand geschaut und als dann richtig produziert wurde und die Vorakzessionen der bibliographischen Daten akribisch durchgeführt wurden, da entdeckte man mehr und mehr, dass die Überschneidungen groß waren. Wir merkten z.B. in einem Projekt, dass viele der geplanten Objekte bereits digital in anderen Häusern vorlagen, dadurch schmolz der eigenen geplante Bestand zusammen, ermöglichte uns aber so, ganz anderen Bestand, hier sogar Zeitungsbestand, für diesen Fachbereich heranzuziehen, der in anderen Häusern nicht so vollständig vorhanden war.

Wir haben natürlich auch Auswahlkriterien für Digitalisierungsprojekten und eines dieser Kriterien ist die Unikalität, was ja schon per se eine Doppeldigitalisierung vermeiden hilft. Wir digitalisieren natürlich nicht nur unikales, es ist aber eines der Kriterien, welches die Priorität von Beständen für die Digitalisierung recht hoch setzt.

If-3. Dokumentation, Metadaten, Erschließungsstandards.

If-3.1. Welche Metadaten werden bei der Digitalisierung erfasst?

Unsere Digitalisate werden alle vollständig in unserem Katalog katalogisiert. Wenn es darin nach heutigen Maßstäben schlechte Aufnahmen gibt, dann würden wir diese zum Wohle des Printexemplars und zum Wohle des Digitalisates noch verbessern, wenn es

die großen Datenmengen zulassen. Die vollständigste Erschließung ist immer in unserem Katalogisierungssystem. Eine Auswahl der wichtigsten bibliographischen Daten lässt sich per Knopfdruck in Goobi übernehmen und hier mit weiteren Daten anreichern. Ein Datum, was wir in Goobi anreichern, ist z.B. die digitale Kollektion. Hier haben wir Digitalisate nach fachlichen Gesichtspunkten zusammengefasst. Ein weiteres Merkmal hier wäre die digitale Serie, früher der Gesamttitel der Sekundärausgabe, hier kreieren wir für solche Projektzusammenhänge einen „Titel“, der dann auch in die Metadaten übernommen wird, aber eigentlich kein bibliographisches Datum ist. Hinzu kommt auch eine URL sowie eine Anreicherung der Strukturdaten in Goobi. Seit einiger Zeit hat sich die Staatsbibliothek zudem nicht nur auf urheberrechtsfreies Material kapriziert, sondern will auch sehen, wie man mit rechtebehaftetem Material zurechtkommt. Vor dem Hintergrund der vergriffenen Werke, die ja jetzt eine gesetzliche Regelung erfahren haben, möchten wir da auch mitmachen, wenn auch nicht in großem Stil, da das Aufgabe anderer Bibliotheken ist. Aber wir haben mehr und mehr Werke, die man uns überlässt, die nicht rechtfrei sind, wo aber Rechteverhandlungen stattfinden und das muss natürlich in irgendeiner Weise dokumentiert werden. Rechtstatements sind dabei weder für die Katalogdaten noch für die Übernahme in Goobi ein einfaches Thema.

Was noch hinzu kommt wären in den Images TIFF-Metadaten: Der Erzeuger (Staatsbibliothek), der eingesetzte Scanner und die verwendete Software zum Erzeugen sowie Bearbeiten der Scans. Hinzu kommen Informationen zum eingebetteten Farbprofil, zum Copyright sowie teilweise eine kurze Beschreibung. Natürlich könnte man das perspektivisch noch ausbauen, aber erstmal ist es ein guter Standard.

If-3.2. Stehen alle erfassten Metadaten dem Nutzer frei zur Verfügung?

- Wenn nein: Welche nicht?

Ja, vollkommen. So ist unser Verständnis, dass sämtliche Daten vollkommen uneingeschränkt nachgenutzt werden können.

Im übrigen ist auch das, was wir im Digitalisierungszentrum digitalisieren und in die digitale Bibliothek überführen gemeinfrei und seit einiger Zeit auch in den Digitalisaten als gemeinfrei gekennzeichnet.

If-3.3. Nach welchen Metadatenschemata werden die Digitalisate beschrieben?

Nach METS/MODS.

If-3.4. Nach welchen Erschließungsstandards werden die Daten beschrieben? Kommen Normdaten zum Einsatz?

Ja, Normdaten kommen unbedingt zum Einsatz, aber eben nur im Katalogisierungssystem. Die Normdaten, ebenso die Verlinkungen, nehmen wir aus technischen Gründen nicht in die Metadaten mit. Das müssten wir händisch übertragen und das schaffen wir einfach nicht. Wir träumen davon, in einer neuen Version die Daten neu zu prozessieren und dann zumindest die ID aus dem CBS für die GND zu übernehmen. Im Katalogisierungssystem wird aber ordentlich katalogisiert und da ist das Level unser normales Katalogisierungslevel inklusive Normdaten.

If-3.5. Wird wenn möglich auch der Volltext erschlossen?

- Wenn ja, in welchem Format wird dieser gespeichert?

Wenn möglich kann man vollherzlich mit ja beantworten. Es kommt vor, aber es ist bisher noch nicht so, dass wir eine OCR-Strecke betreiben oder systematisch mit einem Dienstleister zusammenarbeiten. Wann immer es bisher möglich war wurden aber einzelne Projekte auch im größeren Umfang volltexterschlossen. Das Thema OCR ist für uns aktuell, wie wir dahin kommen, dass wir zeitnah nach der Digitalisierung in den Sammlungen Volltext präsentieren können. Wir können sie zwar schon präsentieren, aber noch nicht systematisch erzeugen.

Welches Format dabei genutzt wird, lässt sich noch nicht sagen. Bisher speichern wir in ALTO-XML, einige Handschriften vielleicht in TEI. FineReader wird zwar auch vorgehalten, aber in den Metadaten findet man bisher wie gesagt ALTO.

If-4. Veröffentlichung.

If-4.1. Unter welchen Lizenzen sind die Digitalisate zugänglich?

- Welche Lizenz wird favorisiert? Warum?

Im Standardfall erfolgt die Kennzeichnung als gemeinfrei über Public Domain. Hinzu kommen drei weitere Fälle: alle Creative Commons Varianten, ein In-Copyright-Hinweis sowie konkret verhandelte Lizenzen, die benannt und in den Akten dokumentiert werden. Wir führen aber auch Verhandlungen, um Public Domain zu erreichen. Die Kollegen sind schon darum bemüht es so einrichten zu können, dass man uns Public Domain einräumt. Public Domain hat den Vorteil, dass jeder Nutzer das machen kann, wozu er das Digitalisat braucht. Lange Zeit hatten wir eine Creative Commons Lizenz, die jedoch eine kommerzielle Nutzung ausgeschlossen hat, was dazu führte, dass Dienste wie Wikipedia diese Digitalisate nicht verwenden durfte, was nicht im Sinne der Bibliothek ist. Man muss dass im Gesamtgefüge der Stiftung betrachten, in der ja auch Museen sind, in denen diese Diskussion ganz anders geführt wird. Während Bibliotheken bei der Frage relativ offen sind und die Meinung vertreten, dass die Digitalisierung nicht die erforderliche Schöpfungshöhe mit sich bringt, um bei einem gemeinfreien Werk einen neuen Schutz zu rechtfertigen, sieht man dies in Museen teilweise anders. Hier sind die Aufnahmeverfahren andere und individueller, aber es bestehen auch andere Interessen als in Bibliotheken. Stiftungsintern konnten wir uns so aufteilen, dass die Bibliothek so arbeitet, wie sie es für richtig hält und die Museen ihren Weg gehen.

Wenn man den Aspekt der Metadaten betrachtet, dann ist das auch anders als in Museen. Während unsere Metadaten und Katalogdaten frei sind, kann in Museen eine Autorenschaft für die Metadaten beansprucht werden. Hier vertritt man die Meinung, dass ein Objekt ohne die Metadaten überhaupt nicht zu verstehen ist, während die Metadaten in Bibliotheken traditionell ein Mittel sind, um die Objekte aufzufinden.

Zwischenfrage: Mit der Public Domain verschwindet ja auch die Namensnennung. Ist dies durch die Staatsbibliothek bewusst gewählt?

Da gab es auch diverse Diskussionen. Die Marke Staatsbibliothek verschwindet ja dadurch. Aber alle möglichen Markierungen, wie z.B. Wasserzeichen, sind sinnlos. Alles, was wir reinrechnen, kann jeder wieder rausrechnen, also können wir es gleich sein lassen.

Außerdem ist es ja gemeingut, was wir produzieren. Hier gibt es aber auch ganz unterschiedliche Standpunkte, die teilweise pragmatisch, teilweise prinzipiell sind und viel mit der persönlichen Identifikation mit der Arbeit zu tun haben.

[Herr Siegmann] Ich persönlich finde es durchaus aus institutioneller Sicht an manchen Stellen hilfreich, Images zu kennzeichnen. Nicht, um irgendwelche Rechte zu sichern, sondern da es ein gutes Werbemittel ist. Die französische Gallica macht es z.B. so, indem sich unter dem Bild ein einfacher weißer Streifen mit dem Hinweis Gallica.fr befindet. Und ich habe schon vielfach Präsentationen gesehen, in denen dieser Streifen nicht entfernt worden ist und man sah daher gleich, dass das Objekt aus der Gallica stammt. Ich finde das persönlich nett, aber das ist nicht die Hausleitlinie.

If-5. Interoperabilität und Schnittstellen, Portale.

If-5.1. Können die Metadaten über maschinenlesbare Schnittstellen geharvestet werden?

- Wenn ja, welche Schnittstellen werden angeboten?

Ja, einmal selbstverständlich über das Katalogisierungssystem über die SRU-Schnittstelle des GBV. Wir selber bieten unsere Daten auch über die OAI-Schnittstelle an, dazu muss man aber sagen, dass wir die OAI-Schnittstelle noch nicht bewerben. Wer damit umgehen kann, findet die Schnittstelle aber und kann sie gut nutzen. IIF ist als Schnittstelle auch schon in die Digitalen Sammlungen integriert, wird aber auch noch nicht beworben.

If-5.2. Werden die Digitalisate auch in andere Angebote oder Portale (z.B. die Deutsche Digitale Bibliothek/Europeana) eingespielt?

- Wenn ja, in welche Portale spielen Sie die Daten ein? Warum gerade in diese Portale?

- Wenn nein, warum nicht?

Wir haben mit der Digitalisierung von Drucken angefangen und diese sind an das ZVDD übertragen. Das funktionierte und funktioniert noch immer gut.

Mit dem Europeana-Projekt zum Ersten Weltkrieg ging die Verpflichtung einher, die Daten bis zum Projektabschluss in der Europeana zu präsentieren. Die einfache Antwort ist daher ja, die Daten werden bereitgestellt und können über OAI geharvestet und präsentiert werden. In der Praxis ist dies jedoch extrem komplex. Die Metadaten sind nicht über alle Objekte vollständig einheitlich, einige Daten sind mal vorhanden, mal nicht. Besonders bei mehrbändigen Werken etc. wird dies dann besonders komplex und führt in der Praxis oft nicht ohne weiteres zum gewünschten Ergebnis. Nichtsdestotrotz gibt es einen (nicht vollständigen) Datenabzug der StaBi in der Europeana und der DDB. In der Theorie sollte es so sein, dass die DDB regelmäßig unsere Daten über OAI harvestet, aktualisiert, präsentiert und sie dann über ein Aggregatorennetzwerk an die Europeana weitergibt. Das funktioniert aus vielen Gründen nicht so, wie es sollte, was aber nicht unbedingt an der Europeana oder der DDB liegt, sondern es sind einfach extrem komplexe Vorgänge, die auch technisch sehr aufwändig sind. Und dadurch ist die theoretische Wirklichkeit nicht vollständig mit der Praxis deckungsgleich.

Neben den Schwierigkeiten beim Metadatenmanagement gibt es aber auch organisatorische Schwierigkeiten gerade bei der DDB. Die DDB hatte ein Datenmodell

und eine Möglichkeit, Daten in einer Pilotphase aufzunehmen. Nach dem Ende dieser wurden die Weichen umgestellt und man musste sich umorientieren und seit dem stehen wir in der DDB ein bisschen in den Warteschlange. Aber wir haben am Haus einen Kollegen, der sich mit dem Mapping zur DDB beschäftigt. Bei der Europeana sind wir auch in so einer Schleife. Wir hatten zwar einen direkten Zugang zur Europeana, diese war aber auch organisatorisch im Umbruch, da man das Parallelprojekt TEL, The European Library, übernahm, welches eher auf Forschungs- und Nationalbibliotheken ausgerichtet war und die Aufgabe des Harvestings hatte. Wir kamen mit unserer Produktion genau in diese Umbruchphase, konnten die Daten also nicht mehr über TEL harvesten lassen, aber das Harvesting durch die DDB und über diese in die Europeana funktioniert nicht.

Im Grunde sind die Europeana und die DDB vor allem dann nützlich, wenn sie komplexe Metadaten sorgfältig und korrekt abbilden. Dieser Aufwand ist aber kaum zu stemmen, weshalb es vielfach Einrichtungen (nicht wir) gibt, die ein vereinfachtes, hoch standardisiertes Metadaten set (Dublin Core) hinschicken, welches vernünftig angezeigt wird, in dem Funktionen aber nur noch eingeschränkt nützlich sind.

If-7. Organisatorisches.

If-7.1. Welche Fördereinrichtungen (DFG, BMBF, EU, Land Berlin, ...) unterstützen Ihre Einrichtung?

Wir haben für Digitalisierungsprojekte sowohl bei der DFG, dem BMBF und der BKM und der EU Mittel bekommen, über das EFRE-Projekt wurden auch aus dem Land Berlin Gelder zugewiesen.

Das funktioniert dann so, dass es eine Ausschreibung gibt, wir bewerben uns mit geeigneten Inhalten und werden im besten Falle ausgewählt. Das hat oft geklappt, ist aber kein totaler Automatismus.

Die StaBi selber ist eine Bundeseinrichtung, die auf Basis des Stiftungsgesetzes eine Bund-Länder-Finanzierung hat, aus der ein Haushalt entsteht, der die tägliche Arbeit finanziert.

If-7.2. Besteht die Möglichkeit, neue Funktionen in das System der digitalen Angebote zu integrieren?

- Wenn ja, wie oft geschieht dies?

Ja. Eine neue Funktion war z.B., dass die OCR eingebunden und sichtbar gemacht wurde. Aber sowas ist mit viel Aufwand verbunden, da wir es über unsere IT-Abteilung selber machen.

Es gibt dazu eine Arbeitsgruppe, die so etwas sammelt und mitentwickelt, die AG Präsentation. Die Präsentation wird schon kontinuierlich verbessert und es sind immer wieder neue Features hinzu gekommen bzw. Fehler wurden nach und nach verbessert, dies ist aber ein kontinuierlicher Prozess, bei dem man nie an ein Ende kommen wird und was oft länger dauert, als man es sich in einer idealen Welt vorstellen würde.

Zwischenfrage: Kann man ungefähr sagen, wie oft neue Funktionen im System nötig werden bzw. eingespielt werden.

Nötig werden und umgesetzt werden sind ja zweierlei Dinge. Ich weiß, dass die AG sich eine ganze Menge wünscht und schöne Ideen zur Vorstellung hat, aber da die Kapazitäten begrenzt sind, lassen sich diese Ideen nur langsam umsetzen. Es sind wie gesagt eher kontinuierliche Verbesserungen.

Eine deutliche Verbesserung war z.B. die Zoom-Funktion in der Darstellung von Karten. Die Funktion war von der Kartenabteilung lange Zeit gewünscht, aber die Umsetzung dauerte dann trotzdem eine ganze Weile. Auch andere Dinge wie die IIIF-Integration, die durch einen neuen Viewer unterstützt wird, der neue PDF-Download, die Integration von Volltexten oder allgemeine Änderungen an der Darstellung wie z.B. Änderung der Reihenfolge der Icons lassen sich nennen.

Was in dem Kontext interessant ist, ist, dass man die offiziell freigegebene Version unter digital.staatsbibliothek-berlin.de hat, aber auch für jeden Nutzer freigegeben eine Testversion unter digital-beta.staatsbibliothek-berlin.de, ganz bewusst als Spielwiese zum Testen. Hier garantiert man zwar keine volle Funktionalität, aber die neuen Features sind schon frühzeitig integriert und sichtbar.

If-7.3. Gibt es Erhebungen darüber, welche Nutzergruppen Ihre digitalen Angebote (am meisten) nutzen?

Ja gibt es, über das Tool Matomo, ehemals Piwik. Dies sind aber nur systematische statistische Auswertungen aller Online-Angebote der StaBi. Dabei zeigt sich, dass die Digitalen Sammlungen neben dem Online-Katalog das größte genutzte Angebot sind und die Nutzerzahlen hier kontinuierlich wachsen. Es gibt jedoch keine Erhebungen im Sinne einer konkreten Nutzergruppenzuordnung, sondern nur, was man eben über solche Webstatistiken erfassen kann: Zu welchen Zeiten kommen Nutzer aus welchen Ländern bzw. mit welchen IP-Adressen und wie lange bleiben diese.

Es gab auch mal vor einigen Jahren eine Nutzerumfrage, hier wurde abgefragt, wie die Nutzer im Katalog suchen, ob sie eher digitale oder analoge Medien nutzen und eben auch, welche fachliche Nutzerschaft das war. Aber das war wie gesagt punktuell und nicht flächendeckend. Über die normale Auswertung erheben wir daher keine Nutzergruppen.

I-1. Auswahl.

I-1.1. Wie sieht der Auswahlprozess von Objekten für Digitalisierungsprojekte aus?

- *In wie fern sind Kuratoren/Sammlungsleiter und Konservatoren beteiligt?*
- *Welche Rolle spielen Schäden an den Objekten, z.B. besonders für oder gegen eine Digitalisierung?*
- *Fällt die Entscheidung vor allem durch Mittelgeber oder nach eigenen Bedürfnissen?*
- *In wie Weit werden die Nutzerbedürfnisse berücksichtigt bzw. die Nutzer eingebunden?*

Fangen wir mal allgemein an. Es gibt den sog. Masterplan Digitalisierung. Dies ist ein auf einer Liste von Auswahlkriterien beruhendes Ranking aller möglichen Digitalisierungsprojekte. Diese beziehen sich auf den hier im Haus vorhandenen Bestand

und wurden vor über fünf Jahren in allen Abteilungen systematisch abgefragt. Die Kriterien dabei sind Unikalität, Fragilität, Relevanz für Wissenschaft und Forschung, FID-Zugehörigkeit etc. Diese Liste ist nicht statisch, da kommen immer neue Projekte hinzu oder einige rutschen in der Priorität weiter nach unten. Das passiert vor allem auf Hinweis der Sammlungsleiter, der bestandsführenden Abteilungen oder der Fachreferenten. Wichtig werden können Bestände dagegen zu bestimmten historischen Ereignissen. Solche Ereignisse haben schon einen Einfluss auf die Auswahl, sind aber nicht das allein ausschlaggebende Kriterium. Die Frage, wie man diese Kriterien gegeneinander gewichtet, stellt sich jedes mal neu. Man kann nicht objektiv sagen, ob die Digitalisierung von historischen Karten der Stadt Berlin gegenüber mittelalterlichen Handschriften wichtiger ist. Das ist zu unterschiedlich im Inhalt, als dass man objektiv sagen könnte, eins ist allgemein wichtiger als das andere. Sowas muss man immer im Einzelfall entscheiden. Und daher gibt es dann noch eine zweite Ebene an pragmatischen Kriterien, die da mit reinspielt. Dies ist natürlich auch, ob das Projekt gerade drittmittelgefördert ist, also die Möglichkeit besteht, einen Bestand, den man eh digitalisieren lassen möchte, nun auch finanzieren zu können. Aber auch, ob wir gerade geeignete Scanner haben. Es hilft nichts, wenn man großformatige Karten mit höchster Priorität digitalisieren möchte, aber nur einen Großformatscanner hat, der gerade voll ausgelastet ist. Solche Aspekte fließen auch mit ein.

Dazu gibt es eine Arbeitsgruppe, das sog. Content Board, welche sich mit den neuen Vorschlägen für Digitalisierungsvorhaben befasst, versucht, diese in das Ranking einzusortieren und dann Empfehlungen für den Lenkungsausschuss bzw. die Generaldirektion ausspricht, welche Projekte für die Digitalisierung vorgeschlagen werden. Dass sich das in der Realität nicht idealtypisch abbilden lässt ist klar, aber das ist im Grunde das Verfahren, wie wir es praktizieren.

Konservatorische Aspekte spielen auf jeden Fall mit in die Entscheidung hinein, da sichergestellt werden muss, dass ein Bestand schadenfrei digitalisierbar ist. Das ist unter anderem dadurch sichergestellt, dass bei Projekten, die in den Bereich des realisierbaren rücken, unsere Restauratorinnen durch die Sammlung gehen und sich die Bestände ansehen. Zum einen ist das wichtig für eine Aufwandsabschätzung, aber auch um zu wissen, wie weit sich die Bände öffnen lassen, welche Scanner wir einsetzen können, wie sich der zeitliche Ablauf gestaltet etc. und auch um sicherzustellen, dass man eine restauratorische Notwendigkeit im Vorfeld erkennt und diese entsprechend berücksichtigen kann. Und dann gibt es natürlich auch noch solche Glücksfälle, dass Bände, die ohnehin gerade in der Restaurierung aufgebunden sind, bspw. als einzelne Seiten digitalisiert werden können und man so wirklich das gesamte Objekt vollständig erfassen kann. Sowas versuchen wir natürlich auch zu nutzen.

Neben diesem Auswahlverfahren gibt es natürlich noch das DoD-Programm. Hier bestellt ein Nutzer ein Buch, wir prüfen, ob das gemeinfrei und digitalisierbar ist, der Nutzer seine Rechnung bezahlt und dann digitalisieren wir. Hier haben wir auch einen Sonderfall, ein DoD-Großauftrag aus der Humboldt-Universität. Hier wurden jetzt schon seit Jahren immer wieder DoD-Aufträge für die Digitalisierung des Briefnachlasses der Gebrüder Grimm in Auftrag gegeben, was so nach und nach dazu führen wird, dass der Nachlass vollständig digitalisiert worden ist. Dadurch erfüllen sich halt auch in gewisser Weise

Nutzerbedürfnisse. Wir suchen aber noch nach einem Weg, wie man auf eine systematisierte Art und Weise berechnete Forschungs- bzw. wissenschaftliche Bedürfnisse im Vorfeld erkennen kann, um die Priorisierung der Digitalisierungsprojekte dementsprechend anzupassen. Wir sind in der glücklichen Lage, dass wir unseren Bestand selbst digitalisieren. Teilweise gibt es auch Anfragen, wo der Bestand der StaBi den Bestand einer anderen Einrichtung z.B. für ein Forschungsprojekt gut ergänzen würde. In früherer Zeit haben wir unsere Bestände dann übergeben, die Digitalisate wurden dort in gleicher Qualität wie bei uns produziert, dort präsentiert und dauerhaft gesichert. Da wir jedoch nicht genügend Kapazitäten hatten, diese Digitalisate dann in unser System einzuspielen, mussten wir zu der Maßnahme greifen, diese Fremddigitalisate nur nachzuweisen. Rechtlich könnten wir diese Digitalisate zwar natürlich bekommen und selber bereitstellen, aus Kapazitätsgründen verzichtete man hier jedoch darauf, da man sie in die gesamte Infrastruktur einbauen müsste.

Zwischenfrage: Heute würde man keine Bestände mehr rausgeben?

Immer weniger gerne, das liegt aber an der zunehmenden Leistungsfähigkeit unseres Digitalisierungszentrums. Es kommen zwar immer wieder Anfragen zu speziellen Interessen, aber hier können wir nicht singulären Einzelimpulsen folgen, sondern müssen auf das DoD-Programm verweisen. Ich habe letztes auf einer Tagung einen spannenden Ansatz aus der dänischen Nationalbibliothek gehört, welche auch ein DoD-Angebot hat. Hier hat jeder Nutzer ein Kontingent an x freien Seiten, welche man ohne irgendein Entgelt bekommen kann. So könnte man zwar bestimmt auch die Wünsche der Nutzer abbilden, was auch wieder anderen Nutzern zu gute käme, das stellt dann aber wieder Fragen an die Ressourcen im Haus bzw. den Fachabteilungen, die für die Digitalisierung benötigt werden.

I-1.2. Wie sieht der Umgang mit Dubletten aus: Werden beide Exemplare, z.B. eines Buches, digitalisiert, wird ein beispielhaftes Objekt aus beiden Dubletten erstellt oder wird sich für eines entschieden?

Das ist auch lange diskutiert worden, ob wie schönere Exemplare erzeugen, als in unserem Buchbestand vorhanden. Nach längerer Diskussion haben wir uns darauf geeinigt, dass Exemplar Exemplar bleibt. Wir streben an, bei mehreren Exemplaren natürlich zu prüfen, welches das vollständigste ist und dann dieses auszuwählen. Zudem vermeiden wir Doppeldigitalisierungen, es sei denn, es gibt Exemplarspezifika. Z.B. haben wir einen Druck, den wir in einem Digitalisierungsprojekt digitalisiert haben und in der Handschriftenabteilung haben wir einen Band, gleiche Auflage, gleiche Ausgabe, allerdings mit Marginalien einer bedeutenden Persönlichkeit, diesen haben wir dann natürlich auch digitalisiert. Das sind aber Ausnahmen, noch nicht mal im VD18 würde mir einfallen, dass wir da mehrere Exemplare eines Bestandes digitalisiert haben.

I-2. Digitalisierung.

I-2.1. Erfolgt die Digitalisierung immer vom Original oder wenn vorhanden auch vom Film/Microfiche?

Wenn ein Original vorhanden ist, dann immer vom Original, mit Ausnahme von Zeitungen. Das ist eine Entscheidung die Zeitungen betreffend, da diese oft in einem konservatorisch schwierigen Zustand sind, also sehr fragiles Papier haben, oft auch stark in den Falz hineingebunden sind und ohnehin schon verfilmt wurden. Hinzu kommt, dass das Digitalisierungsverfahren vom Film einfach viel schneller ist und der Qualitäts- und Erkenntnisgewinn durch eine Digitalisierung vom Original dagegen nicht ins Gewicht fällt. Bei historischen Zeitungen geht es im wesentlichen um die Textinformation und die lässt sich vom Film in der Regel genauso gut aufnehmen wie vom Original, aber man hat hier einen viel größeren Durchsatz, so dass wir uns hier für die Filmdigitalisierung entschieden haben. Dann gibt es vereinzelt Aufträge zur Digitalisierung vom Microfilm, das sind Bestände, die im Original nicht mehr verfügbar oder die für die Benutzung gesperrt sind. Diese wären dann nicht vom Original digitalisierbar. Damals hatte man eine Sicherheitsverfilmung durchgeführt, eben um die Bestände zu retten. Inzwischen machen wir auch noch eine Sicherungsdigitalisierung, um Dinge, die vor unsern Augen verfallen, auch digital zu retten. Die Originale werden dann nicht mehr herausgegeben und von den Konservatoren so verpackt, dass sie zwar noch existieren, aber eben nicht mehr in Benutzungssäle kommen.

I-2.2. Als Qualitätsstandard werden oft die DFG-Praxisregeln zur Digitalisierung genannt. Wie relevant sind diese für die Digitalisierung in ihrer Einrichtung?

Die DFG-Standards sind sehr wichtig. Sie sind in allen Digitalisierungsprojekten die Leitlinie und definieren – auch den nicht DFG-geförderten Projekten – den Mindeststandard, den wir in bestimmten Fällen übertreffen, aber nie unterschreiten.

I-2.3. Welche Maßnahmen zur Qualitätssicherung werden getroffen?

Technisch haben wir wie gesagt versucht, uns, was die Abbildungsqualität der Scanner anbelangt, mehr und mehr an Standards wie Metamorfoze oder ISO 19264-1 auszurichten, haben seit 2017 nun auch Geräte, die das erfüllen. Dadurch können wir sicherstellen, dass die Scanner bestmögliche Abbildungsqualität liefern. Das wird durch Wartung, regelmäßige Reparaturen etc. sichergestellt.

Für die Qualitätskontrolle im direkten Produktionsprozess haben wir Mitarbeiter, die alle Images bzw. Vorgänge, die an den Scannern erzeugt werden, nochmal darauf kontrollieren, ob das Bild scharf, das Objekt vollständig ist, die Images gerade sind, ob ein ColorChecker vorhanden ist etc. Zu jedem Vorgang gibt es eine Farbtafel, einen Auflösungsmire, ein Hinweis, wer das Digitalisat erzeugt hat sowie ein Zentimetermaß, so dass ein Nutzer die Originalgröße der Vorlage möglichst rekonstruieren kann.

Darüber hinaus orientieren wir uns bei der Präsentation und Metadatenerfassung an marktüblichen Standards und setzen weitestgehend Open-Source-Software ein. Zudem verwenden wir Dateiformate, deren Format offengelegt ist, so dass man sich im schlimmsten Fall einen Viewer selber zusammenbauen kann und die Daten nicht verloren gehen. Alle Digitalisate werden zudem gespiegelt, sind also mehrfach vorhanden, und auch auf einem weiteren Backup gesichert.

I-2.4. Werden Daten im Nachhinein noch verändert?

- Wenn ja, findet eine Erfassung der Veränderungshistorie statt?

Ja, wir verändern. Manchmal fällt uns hinterher auf, dass eine Seite oder die ganze Lage fehlt, mal ist auch das falsche Buch gescannt, welches dann aus der Präsentation wieder raus muss. Wir verbessern unsere Digitalisate, aber eine Versionierung, also eine wirkliche Dokumentation vor allem auch der Metadaten, die dazugehört, das haben wir leider noch nicht.

Was nicht erfolgt sind Bildmanipulationen im Sinne von Photoshop, also dass man z.B. die Bildhelligkeit im Nachhinein erhöht. Das Master wird so gespeichert, wie es vom Scanner kommt. Es wird für die Präsentation ein Derivat vom Master erzeugt, dass eine JPEG-komprimierte TIFF-Datei mit 80% Kompressionsgrad präsentiert wird. Das hat vor allem auch den Grund, dass die Dateien so handhabbar und downloadbar sind, da sie so weniger als ein Viertel der Ursprungsgröße haben.

I-3. Dokumentation, Metadaten, Erschließungsstandards.

I-3.1. Welche Informationen zum Kontext der Digitalisierung (Sammlungsbereich, Provenienz, Digitalisierungsprojekt, ...) werden dokumentiert und verfügbar gemacht?

Das kommt auf das Bestandssegment an. Im neueren Bestand, der beginnt bei uns so ab 1850, ist es so, dass wir Provenienzvermerke nicht routinemäßig machen, außer, wenn wir Sammlungszusammenhänge darstellen. Das ist keine echte Provenienz, also durch wessen Hände das Buch gegangen ist, sondern wo zum jetzigen Zeitpunkt festgestellt wird, welchen Sammlungszusammenhang das Werk hat.

In älteren Beständen betreiben wir auch direkt Provenienzforschung, das ist sehr aufwändig und wird in unserem Provenienz-Wiki festgehalten. Diese Daten werden aber nicht als Metadaten an das digitale Projekt weitergereicht, sondern verblieben im Katalogisierungssystem.

Jedes Vorhaben wird dabei mit einem Projekttitel versehen.

I-3.2. Durch die Digitalisierung gehen einige Informationen wie z.B. die Materialität der Quelle in der Nutzung verloren. Können umfangreiche Metadaten hier einen gewissen Ausgleich schaffen?

Oh ja, das denken wir ganz bestimmt. Gerade für die älteren Schriften, wo auch der Einband oder das Papier eine ganz andere Aussagekraft und einen eigenen Wert hat als bei moderner Literatur. Da gibt es durchaus auch Nachweisinstrumente, wir haben so z.B. eine Einbanddatenbank, in der die Kollegen solche Daten festhalten. Im Katalogisierungssystem ist dies leider nicht sichtbar, aber es wäre schön, wenn diese Daten zusammenfänden. Mehr Metadaten würden da ganz sicher einen Ausgleich schaffen.

Zwischenfrage: Vielleicht auch Metadaten, an die man bisher nicht denkt?

Ganz sicher. Ich konnte mir mal die Metadaten der Einbanddatenbank anschauen, welche Fakten die Kollegen hier beschreiben. Das ist wirklich eine eigene Datenbank wert, nicht nur eine Kategorie irgendwo in der Titelaufnahme, sondern das ist eine sehr deutliche Beschreibung. Ich würde mir daher eine Verlinkung zwischen der bibliographischen

Beschreibung und dieser ausführlichen Beschreibung des Objektes wünschen. Und das ist nur der Einband. Für die Buntpapiere, die Vorsatzpapiere, die in ganz besonderer Weise gestaltet wurden, haben wir auch eine kleine hausinterne Datenbank. Das wäre aber auch ganz interessant, da es direkt an unsere Objekte gebunden ist und man sagen könnte, dieses Buntpapier findet man in einem Band der Staatsbibliothek der so und so heißt. Das wäre interessant, Informationen zum Papier etc. sowieso. Wir haben sonst noch eine Mitarbeit an einem Digitalisierungsprojekt an der WZIS Datenbank aus Baden-Württemberg. Die weist Wasserzeichen nach. Hier wurden mit speziellen Kameras die Wasserzeichen aufgenommen und in dieser Datenbank speziell beschrieben.

Zwischenfrage: Und denken Sie, dass Digitalisate mit sehr guten Metadaten Originale ersetzen können?

[Frau Müller] Ich erinnere mich an einen Satz, als wir mit der Digitalisierung beginnen sollten. Da sagten die Kollegen, die das für eine Zumutung für unsere Bestände hielten: „Ach, und dann schmeißen wir unsere Original hinterher weg, ja?“ Also ich denke, es ist für uns Bibliothekare kaum vorstellbar, dass wir unsere Bücher wegschmeißen. Die Originale ersetzen? Hier kommt es natürlich darauf an, um welche Originale es sich handelt. Bei alten Drucken wissen wir, dass es da große Unterschiede geben kann und dass es da eben auch ein wirkliches Forschungsfeld gibt, wie welche Exemplare unter welchen Umständen, auch im Prozess der Herstellung, ausgesehen haben. Bei neueren Sachen, so ab der zweiten Hälfte des 19. Jahrhunderts, kann man sicher drüber streiten. Oder eine andere Idee verfolgen, die es auch gerade im bibliothekarischen Umfeld im Bereich der Digitalisierung gab, dass man nur noch ein Exemplar aufhebt, z.B. in der Bibliothek, die zuerst digitalisiert hat. Diese hebt dann auf, die anderen können wegschmeißen. Solche Ideen gab es zwar auch, aber wirklich Bibliotheken ausräumen? Wohl eher nicht.

[Herr Siegmann] Ich sehe das immer von einer etwas anderen Seite. Ich denke, man kann durch die Digitalisierung massive Zusatznutzen schaffen und ganz viele Nutzungszwecke von Büchern vereinfachen. Der Standard-Bibliotheksnutzer interessiert sich in der Regel ja erstmal für den Inhalt. Also den Text, den man in vielfältigster Form wiedergeben kann, wo die Digitalisierung, gerade etwas weiter gedacht, zusätzlichen Nutzen bietet. Man kann die Schrift vergrößern und verkleinern, kann sich den Text vorlesen oder übersetzen lassen, man muss nicht hier her kommen, um den Text zu lesen, etc. Also diese ganzen Zusatznutzen, die überhaupt erst möglich werden durch die Umwandlung dieses graphischen Images in einen Text und die Bearbeitungsformen, die PCs und Informatik bieten.

Auch die ästhetische Seite ist interessant. Man kann so überhaupt erstmal Zugang schaffen zu wirklichen Kunstschatzen, von denen wir ja auch nicht wenige haben. Bücher, die über Jahrhunderte im Magazin standen, geschützt von Torwächtern, die eine wie auch immer geartete Berechtigung abgefragt haben, bevor man sowas überhaupt zu sehen bekommen hat. Im Grunde ist das also ein Demokratisierungsprozess. Jeder, der sich dafür interessiert kommt da ran und kann sich das angucken. Man kann so auch viel von dem ästhetischen Wert weitergeben. Egal ob das jetzt Leonardo da Vinci ist oder irgendeine mittelalterliche Handschrift. Das ist auch ein totaler Zusatznutzen, den das bringt. Ich glaube nicht, dass es rechtfertigt, das Medium an sich deshalb auszusortieren.

Es bleiben natürlich bestimmte Informationen, die halt auch medientypisch sind, immer dem Digitalisat vorenthalten. Die kann man einfach nicht so transportieren. Vielleicht geht es in Zukunft besser, wenn man z.B. ein ganzes Buch als 3D-Objekt aufnimmt. Auch wenn man an Bücher denkt, die man z.B. nicht mehr öffnen kann, dazu gibt es ja auch ganz spannende Verfahren. Letztens wurde auf einem Vortrag vorgestellt, wie man bei einer Rolle, die man aus konservatorischen Gründen nicht mehr ausrollen konnte, mit Hilfe von computertomographischen Verfahren die eisenhaltige Tinte erkennen und so ein digitales Abbild des aufgerollten Originals erstellen konnte. Solche Verfahren sind natürlich super spannend, aber natürlich würde ich solche Originale trotzdem niemals wegschmeißen.

Wer weiß, wo das zukünftige Forschungs- und Erkenntnisinteresse liegt. Solange ich die Möglichkeit habe, das Original zu bewahren, sollte man es tun.

I-4. Veröffentlichung.

I-4.1. Werden alle digitalisierten Objekte (soweit rechtlich möglich) auch in Ihren digitalen Angeboten zur Verfügung gestellt? Sind hier alle angebotenen Digitalisate frei zugänglich?

Wie schon gesagt, soweit rechtlich möglich ist bei uns alles frei zugänglich.

I-4.2. Werden die Digitalisate in der höchstmöglichen Qualität (Master) in der digitalen Sammlung Ihrer Einrichtung angeboten?

- Wenn nein, warum nicht, worin unterscheiden sich die angebotenen Digitalisate vom Master?

- Wenn nein, was ist nötig, um die Master-Dateien zu erhalten?

Aus pragmatischen Gründen bezüglich Übertragungsgeschwindigkeit etc. wird eine JPEG-komprimierte TIFF-Datei in den Digitalen Sammlungen für die Nutzer zur Verfügung gestellt. Die Master sind nicht standardmäßig für die Nutzer zugänglich, es gibt aber die Möglichkeit, im Rahmen von DoD auf Wunsch die Master-Dateien bzw. extrem hochwertige Digitalisate zu bekommen. Wenn ein Nutzer das wünscht, dann stellen wir das auch zur Verfügung.

Zwischenfrage: Also auch bereits bestehende Digitalisate werden über DoD angeboten?

Das hängt dann von dem Nutzer ab. Wenn das Buch bereits digitalisiert ist und der Nutzer mit dem, was in den Digitalen Sammlungen präsentiert wird, was meiner Meinung nach für 99% aller Anwendungsfälle ausreichend ist, zufrieden ist, dann lädt er sich es aus den Digitalen Sammlungen runter. Wenn der Nutzer aber noch höhere Qualität möchte, dann würde er ein Entgelt entrichten und auch das Master bekommen.

I-4.3. Welche Rolle spielen Zertifikate für Forschungsdatenrepositorien (z.B. Data Seal of Approval oder das DINI-Zertifikat) bei Ihren digitalen Angeboten?

Wir wissen, dass es uns fehlt und wir würden es anstreben wollen. Aber mit Bezug auf die Digitalen Sammlungen sind mir keinerlei Bestrebungen bekannt. Bei anderen digitalen Objekten, z.B. den Born Digitals, und für die Repositorien in bestimmten FID-Zusammenhängen wird es durchaus angestrebt. Aber es hat noch keines unserer

Repositorien, die Angebote des GBV mal ausgenommen, bisher solche Zertifizierung. Aber anstreben würden wir sie.

I-5. Interoperabilität und Schnittstellen, Portale.

I-5.1. Findet eine Verlinkung zwischen Objekten (intern wie extern) statt?

Zwischen den Objekten selbst nur insofern, als dass wir bei der Katalogisierung in den allermeisten Fällen eine Verlinkung zwischen der Beschreibung des analogen und des digitalen Objektes machen. Insofern kann man in unserem Katalog dann zwischen beiden Dingen hin und her schalten und man wird in der jeweiligen Aufnahme auf das andere Objekt hingewiesen. Man sieht dann, dass es eine digitale Ausgabe gibt und umgekehrt. Weitere Verlinkungen würden wir uns wünschen, gerade auch über die gesamte Stiftung Preußischer Kulturbesitz mit anderen Objekten, aber das ist im Augenblick nicht der Fall. Je nachdem wie weit man das fasst haben wir natürlich die Digitalen Kollektionen, die alle Digitalisate zu einem Fachgebiet zusammenfassen, es gibt Projektkennzeichnungen, die teilweise für die Geldgeber wichtig sind, um zu schauen, welche Projekte mit ihrer Förderung realisiert worden sind. Das ist natürlich im weitesten Sinne auch eine Verlinkung zwischen den einzelnen Objekten. Und es gibt natürlich einige Projekte, die haben bestimmte eigene Portale, über die dann dieser Zusammenhang gekennzeichnet wird. Das wäre z.B. für das Haus das E.T.A.-Hoffmann-Portal oder das Alexander-von-Humboldt-Portal. Der Nachlass von Alexander von Humboldt liegt ja zum Teil noch in Krakau und es gab zeitgleich zwei Digitalisierungsprojekte, deren Ergebnisse jetzt gemeinsam auf dem Alexander-von-Humboldt-Portal präsentiert werden. Die Images, die in Krakau gespeichert werden und die Images, die hier gespeichert werden, lassen sich dann gemeinsam betrachten. Die Images liegen zwar nicht auf einem Server, aber es werden eben die Bezüge gesetzt, so dass der Nachlass virtuell rekonstruiert wird. Aber wenn man in Richtung LOD denkt, dann gibt es keine Objekte, die so verknüpft sind.

I-5.2. Werden die Digitalisate von Ihrer Seite auch auf weniger wissenschaftlichen Plattformen wie Wikipedia angeboten?

- kurze Begründung

Nein, werden sie nicht. Es gab zwar mal eine Anfrage dazu im Rahmen der Wikipedian in Residence, diese wurde aber bearbeitet und dann abgelehnt, da wir damals noch mit den CC-Lizenzen gearbeitet hatten. Daher nein, nicht systematisch. Es gibt sicher vereinzelt etwas, aber das stammt dann eher von Nutzern, nicht vom Haus selber.

Zwischenfrage: Und wenn man an Social Media Plattformen denkt. Werden die vom Haus aus bespielt?

Wir haben einen Blog, in dem wir in groben Zügen über wichtige Erwerbungen und vielleicht auch das eine oder andere Digitalisat informieren, aber das ist in Bezug auf Digitalisate alles eher zufällig und nicht systematisch. Vieles erfolgt anlassbezogen, so dass es Mitteilungen oder Vorstellungen zu entsprechenden Digitalisaten gibt. Die Humboldt-Tagebücher waren hier z.B. sehr prominent, da wurde dann auch auf die entsprechenden Digitalisate hingewiesen, z.B. mit einem schönen Bild von dem Äffchen

von Humboldt. Wenn wir Digitalisate über Social Media publizieren, dann also nur anlassbezogen, sonst nicht.

Im Digi-Zentrum hatten wir mal den Ansatz, wenn wir was besonderes/interessantes da haben, dann über Social Media darauf hinzuweisen, haben dabei aber gemerkt, dass das in so einem großen Haus recht aufwändig ist. Im Digitalisierungszentrum sind wir aber einfach nicht die Experten dafür, dass wir fachlich etwas gehaltvolles über die Objekte schreiben könnten. Das heißt, man bräuchte dafür jemanden aus der Abteilung und hier eine engere Zusammenarbeit. Das wäre sicher im Sinne der Öffentlichkeitsarbeit sinnvoll und extrem wirkungsvoll, passiert aber leider aufgrund des hohen Aufwands nicht.

I-6. Speicherung, Pflege und Archivierung.

I-6.1. Gibt es spezielle Strategien (z.B. regelmäßige Migration, Konvertieren, ...) zur Langzeitarchivierung der Daten?

Wir sind beide keine ITler, aber haben so weit Einblick in das Geschehen der IT-Abteilung, dass wir doch was sagen können. Die Kollegen dort geben sich natürlich alle Mühe, dass es jede Art von Sicherung gibt, aber eine wirkliche Langzeitarchivierung, also eine Archivierungsinfrastruktur haben wir noch nicht. Aber im Moment haben wir eine Kollegin im Haus, die sich mit der Planung einer solchen Infrastruktur beschäftigt. Und bis dahin tun wir alles, um unsere Digitalisate so sicher wie möglich unterzubringen.

Einige Beispiele wären gespiegelte und lokal getrennte Speicherorte, einmal im Haus Unter den Linden und im Haus Potsdamer Straße, zum Teil auch in Friedrichshagen, in denen die Images liegen. Selbst wenn eines der Häuser zerstört wird, so gibt es immer noch eine Kopie. Es gibt ein Langzeitbackup und eine tägliche Sicherheitskopie, bezüglich der Nachweise gibt es zudem Ansätze, dass man die Digitalisate nicht über verteilte Systeme speichert, sondern für jedes Digitalisat das wir haben auch mindestens ein sog. Rumpfkatalogisat im zentralen Katalog (CBS) existiert, so dass man zu einem späteren Zeitpunkt, wenn unsere Systeme zusammenwachsen, eben auch die zusammengehörigen Datensätze z.B. aus Goobi und dem Katalog zusammenführen kann. Gerade bei der Digitalisierung von Nachlässen würde es den Workflow wesentlich vereinfachen, direkt aus Kalliope eine Schnittstelle zu Kitodo zu bedienen, führt aber zu dem Problem, dass wir dann verteilte Systeme hätten, was wir nicht wollen.

Also mit den Bordmitteln wird alles mögliche getan, in der Spiegelung liegt ja auch schon eine gewisse Sicherheit, aber eine Langzeitarchivierungsstrategie gibt es noch nicht.

I-6.2. Wie sieht die Backup-Strategie für die Digitalisate aus?

- In welchen Abständen werden Sicherungskopien durchgeführt?
- Werden redundante Sicherungskopien geführt?
- Werden sie in anderen Formaten als im laufenden System gesichert?

Das haben wir ja eben eigentlich schon beantwortet. Spezielle Backup-Formate werden aber nicht verwendet.

I-6.3. Lohnt sich eine langfristige Archivierung von Digitalisaten (im Bezug auf Aufwand, Finanzen, Verwaltung, ...) oder sollten diese bei Bedarf (beschädigte Daten, veraltete Formate, ...) einfach neu vom Original aufgenommen werden?

[Frau Müller] Wir sprachen vorhin ja über Sicherheitsverfilmung, wo man sich vor dreißig, vierzig Jahren dachte, wir verfilmen alle unsere Bestände, legen diese in super Qualität in den Barbarastollen und dann ist alles gut und heute liegen die Bücher wieder unterm Scanner. Wenn wir jetzt wieder von vorne beginne würden... Man wirft die Digitalisate natürlich nicht einfach weg. Also wir sprechen jetzt nicht von irgendeinem landläufigen Druck aus dem 19. Jahrhundert, den man notfalls gerne nochmal unter den Scanner legen kann, sondern von wertvolleren Materialien. Bei diesen stellen wir uns eigentlich nicht vor, dass die irgendwie nochmal neu behandelt werden.

[Herr Siegmann] Es lohnt sich auf jeden Fall. Aber es ist eigentlich ein sehr komplexes Feld. Wenn man von einem Buch ausgeht, kann man dieses ohne Probleme über 500 Jahre im Regal lagern. Solange es von oben nicht reinregnet, die Luftfeuchtigkeit nicht zu hoch ist und kein Ungeziefer reinkommt, lässt es sich relativ sicher aufbewahren und erhalten. Das ist mit digitalen Images nicht so. Der Datenträger hat eine begrenzte Lebenszeit und muss erneuert werden, das produziert natürlich laufende Kosten. Nichtsdestotrotz sind die Speicherkosten kontinuierlich gesunken und der Aufwand zum Digitalisieren bleibt gleich. Wenn man überlegt, was ein Scanner kostet, was eine Scankraft kostet und was auch das Personal drumherum kostet, dann vergessen wir dabei immer sehr schnell, dass das Buch nicht nur auf den Scanner muss, sondern wir haben den gesamten Workflow, beginnend bei der Katalogisierung, Erstellen der Strukturdaten, Präsentation, usw. Die Aufbewahrung lohnt sich daher auf jeden Fall.

Aber das eigentliche Kernargument, was ich hierbei sehe, ist die Bestandsschonung. Wir versuchen ja, und kriegen das auch ganz gut hin, den Band so wieder aus dem Digitalisierungszentrum rausgehen zu lassen, wie er reingekommen ist, also ohne Beschädigung. Natürlich ist die Digitalisierung eine Belastung für den Band und stellt eine intensive Benutzung dar. Der Band wird geöffnet, so gut es eben geht, man betrachtet jede einzelne Seite, blättert und es gibt in gewissem Maße Stöße, Druck etc., die vom Scanner verursacht werden. Daher versuchen wir eben, in höchstmöglicher Qualität aufzunehmen, damit man diesen Vorgang nur einmal durchführen muss. Auch im DoD-Programm. Wenn ein Benutzer aus einem 250-seitigen Band nur zehn Seiten braucht, dann versuchen wir auch wann immer es geht diesen Band vollständig zu digitalisieren, um eben diese Belastungen einer erneuten Digitalisierung, eines erneuten Transports, etc. zu vermeiden.

[Frau Müller] Wir beide argumentieren gerade aus der Sicht der Bibliothek. Es gibt auch noch einen anderen Aspekt, den wir vorhin schon ein bisschen angesprochen haben. Unsere Digitalisate haben im Moment eine PURL, die wir auch persistent halten. Wir streben aber eine Vergabe für DOI an und damit gewährleisten wir ja auch, dass sie zitierbar sind. Und so haben wir uns eigentlich schon verpflichtet, die Digitalisate zu erhalten und eben präsentierbar vorzuhalten, so dass das Zitat nachvollziehbar ist. Eigentlich verstoßen wir ja schon gegen diese Richtlinien, da wir ja wie gesagt unsere Digitalisate verändern, also z.B. fehlende Seiten einfügen oder ähnliches. Aber wenn wir uns gerade bei den rechtefreien Materialien für eine DOI entschließen und das auch zitiert

wird, dann müssen wir dem auch treu bleiben und sind damit eigentlich alle Verpflichtungen zur Langzeitarchivierung eingegangen.

I-6.4. Sollten bei einer qualitativ besseren Neuaufnahme diese Digitalisate an Stelle der alten Daten treten oder sollte hier ein neuer Datensatz angelegt und mit dem alten verlinkt werden?

Da wir anstreben, eine eindeutige Referenzierung dauerhaft vorzuhalten, bedeutet das eindeutig Versionierung und eindeutig, beide Versionen vorzuhalten. Wobei das in der Praxis, nach der augenblicklichen Vorstellung, heißen würde, dass man in unseren Digitalen Sammlungen nur die aktuelle Version sehen würde, die alten bleiben im Hintergrund erhalten, die DOIs würden also nicht ins Leere führen und blieben so zitierfähig.

Tatsächlich hatten wir neulich den Fall, dass wir ein Werk nochmals im Digi-Zentrum hatten, obwohl es schon digital vorlag. Das Objekt reflektierte aber sehr stark und wir sind jetzt einfach technisch besser, als wir es noch zum Beginn der Digitalisierung bei der ersten Aufnahme des Werkes waren. Die fotografischen Fertigkeiten, mit solchen Materialien umzugehen und sie möglichst originalgetreu wiederzugeben, sind besser geworden. Das betrifft natürlich nur einen ganz kleinen Teil der Digitalisate, aber es gibt sie und in diesem Fall ist wirklich neu digitalisiert worden. Man merkt schon eine technische Weiterentwicklung, die sich bei so ganz speziellen Anforderungen dann schon bemerkbar macht. Diese Fälle bilden aber wie gesagt nur einen ganz geringen Teil.

Aber diese Ära beginnt, dass die Technik besser wird und es uns dann selbst nicht mehr genügt und wir eine bessere Qualität anbieten könnten. Und damit sind wir dann voll dabei, dass wir für Langzeitarchivierung, persistente Adressierung und Versionierung sorgen müssen.

I-6.5. Werden nicht mehr genutzte oder veraltete Digitalisate (z.B. veraltetes Dateiformat) aus dem System gelöscht?

- Wenn ja, unter welchen Bedingungen genau?

Das ist bisher nicht relevant. Wir haben von Anfang an auf Basis der DFG-Standards digitalisiert und die haben sich bisher nicht groß verändert und das können wir alles erfüllen. Diese Fall haben wir daher nicht.

Es gibt hin und wieder die Diskussion, ob es sich perspektivisch lohnt, auf JPEG2000 zu wechseln. In diesem Fall würden wir vermutlich konvertieren. Der große Vorteil von JPEG2000 ist ja, dass die Datei nur ein Sechstel so groß ist wie bei TIFF, der wesentliche Faktor ist also der geringerer Speicherbedarf und die damit erzielbarere höhere Geschwindigkeit. Dies erzielt man natürlich nicht, wenn man die TIFF-Dateien dann irgendwo liegen ließe. Aber das ist bisher nur eine theoretische Frage.

I-7. Organisatorisches.

I-7.1. Liegt ein Datenmanagementplan oder ähnliches zum Umgang mit den Digitalisaten vor?

- Wenn ja, welche Aspekte des Datenmanagements werden beschrieben?

- Wenn ja, wer ist für die Umsetzung verantwortlich?

Das was wir gerade beschrieben haben, also die Abläufe der IT-Abteilung, ist in keinem Plan festgelegt, sondern einfach die Arbeitsabläufe der Kollegen. Aber einen wirklichen Datenmanagementplan gibt es nicht.

I-7.2. Wer ist Ihre Zielgruppe (Fachcommunity, interessierte Öffentlichkeit, andere Einrichtungen, ...)?

Alle und jeder. Wir sind ja nicht an eine Universität oder eine andere Institution gebunden, wie z.B. an die Max-Planck-Gesellschaft oder ähnliches, sondern wir sind wirklich für jeden da. Partiiell, z.B. in den FIDs, sprechen wir durchaus bestimmte Communities an, auf die dann spezielle Angebote, auch in der Digitalisierung, ausgerichtet sind, aber im Grunde ist für jeden alles.

Wir verstehen uns als Universalbibliothek und dementsprechend schwer ist die Eingrenzung auf bestimmte Zielgruppen. Man würde wahrscheinlich Wissenschaft und Forschung immer einen gewissen Vorrang einräumen, aber auch wenn wir uns unsere Nutzer anschauen, die noch persönlich ins Haus kommen, ist das bunt durch die Gesellschaft gemischt und auch der interessierte Laie ist Nutzer unseres Hauses und herzlich willkommen.

I-7.3. Werden bekannte Nutzergruppen bei Entscheidungen (z.B. was als nächstes digitalisiert wird) mit einbezogen?

- Wenn ja, wie findet die Einbeziehung statt?

Grundsätzlich ja, wir sprachen grade die FID-Communities an. Zu diesen pflegen die Fachkollegen spezielle Beziehungen und kennen diese gut, da gibt es durchaus Bitten. Oder auch in anderen Projekten, z.B. das Grimm-Projekt mit der HU oder ähnliches. Sofern Kontakte da sind, spielt es durchaus eine Rolle.

Es ergeben sich natürlich durch die Arbeit in den Fachabteilungen, die ja viel stärker auch auf wissenschaftlichen Konferenzen und in den wissenschaftlichen Kontexten aktiv sind als wir beide, auch Kontakte, aus denen sich dann wieder Projekte ergeben etc. Das sind natürlich auch die Nutzergruppen im weitesten Sinne. Wir haben keine Liste vorliegen, auf der unsere Nutzergruppen priorisiert sind, sondern sie sind einfach schwerer zu greifen und durch unsere unterschiedlichsten Fachabteilungen am Haus auch sehr vielfältig.

I-7.4. Sollten Einrichtungen die Forschungsdaten, die aus ihren Digitalisaten hervorgehen (z.B. Korpora, Textauswertungen, ...), in Form eines Repositoriums sichern und verfügbar machen?

Oh ja, gerne. Es wäre auch schön, wenn uns die Nutzer darüber Bescheid geben würden und wir dann die Chance hätten, die Dinge einfach miteinander zu verknüpfen. Das erinnert an ein Vorgehen und eine Bitte von Bibliotheken im analogen Bereich. Wenn wir unikales Material in der Benutzung zur Verfügung stellen und daraus entsteht ein Zeitschriftenaufsatz oder eine Forschungsarbeit, dann erbitten wir diese im analogen Bereich und die Verfasser sind in der Regel so freundlich, uns diese abzuliefern. Im digitalen Bereich sollte das nicht viel anders sein. Es ist aber ein Wunsch und darauf haben wir kein Recht. Aber es wäre schon toll.

Zwischenfrage: Sollten Bibliotheken dann auch aktiv auf die Nutzer zugehen und aktiv anfragen, ob man aus eigenen Materialien resultierende Forschungsdaten erhält?

Ganz genau. Wir hatten solche einen Ansatz, da ging es aber nur um Katalogdaten, die wir bereitstellen wollten, die Aufgabe wurde aber durch den Verbund übernommen. Hier bestand aber die Absicht, in einem Angebotstext die Bitte zu integrieren, wenn daraus eine Arbeit entsteht, dies doch bitte der Staatsbibliothek mitzuteilen und in irgendeiner Weise zugänglich zu machen. Also grundsätzlich denke ich ja. Wenn wir es denn überhaupt mitkriegen. Gerade bei der OAI-Schnittstelle wissen wir, dass die DDB und die Europeana diese mühsam nutzen, da wir bei diesen Portalen mitmachen wollen. Wer sonst noch die OAI-Schnittstelle nutzt, möchte ich nicht wissen und auch im Haus möchte das niemand wissen. Wir bieten die Daten hier an, die Nutzer können sie einfach verwenden. Dadurch ist es halt ein bisschen schwierig mit denjenigen in Kontakt zu kommen, die die Schnittstelle wirklich nutzen, da sie sich aus technischen Gründen bei uns gar nicht melden müssen.

I-7.5. Abschluss: Welche Aspekte sollten sich Ihrer Meinung nach bei der Digitalisierung in Gedächtniseinrichtungen verbessern?

[Herr Siegmann] Ich würde z.B. Angebote wie die DDB/Europeana sehen. Ich glaube da stecken große Chancen drin, sehe aber noch nicht, dass diese, auch auf Grund der Begrenztheit der Mittel, in dem Maße umgesetzt werden, dass sie das volle Potential entfalten können.

Genauso der Umgang mit Rechten. Die Urheberrechte gibt es und diese müssen natürlich gewahrt bleiben, aber jenseits der urheberrechtlichen Begrenzung, der Umgang mit gemeinfreien Werken. Da glaube ich entstehen noch Potentiale durch gewisse Einstellungswandel in den Einrichtungen, die auch mehr und mehr stattfinden und die dann auch der Wissenschaft und der Allgemeinheit zu Gute kämen, da die Bestände dann auch einfach nutzbarer werden.

Alles andere ist im Grunde eine Frage der Mittel. Natürlich wäre es toll, wenn wir jeden Nutzerauftrag in höchster Geschwindigkeit sofort erfüllen könnten, aber das ist letztlich eine Frage, wie viel Personal und wie viel Geld man zur Verfügung hat.

[Frau Müller] Wir sprechen ja für ein sehr großes Haus und eine sehr große Sammlung. Wir hätten also ganz viel Material, was wir digital zur Verfügung stellen könnten. Und wir müssen uns manchmal die Kritik gefallen lassen, dass wir zwar viel digitalisieren, aber eigentlich anderes gebraucht wird. Das kommt schon aus der Nutzerschaft, gerade aus den Communities unserer FIDs oder dem Forschungsumfeld, und hier wäre mit noch lieber, wenn diese Dinge näher zusammenfinden könnten, so dass man durchaus für einen Forschungsbedarf einen gewissen Grundstock legen kann, wie wir es aus bibliothekarischer Sicht, was wir lesenswert und verbreitenswert halten und auch unter den Bedingungen, die uns unsere Förderinstitutionen eröffnet haben, bereits getan haben. Aber darüber sind wir eigentlich hinaus. Sondern wir sind jetzt an einer Nutzerschaft angekommen, von der ich manchmal mehr wissen müsste, was diese braucht und möchte. Es passiert durchaus, dass einzelne Wissenschaftler Kontakt zu uns suchen. So regte z.B. ein Literaturwissenschaftler an, einen bestimmten Bestandteil zu digitalisieren für konkrete Arbeiten, die er vor hatte. Zwar kam es aus konservatorischen Gründen nicht

dazu, aber in dieser Art wünsche ich mir noch mehr Anregung, um auch ein bisschen zielgerichteter zu arbeiten. Nicht unbedingt und immer, aber eben auch.

Zwischenfrage: Sollten Bibliotheken dann direkt auf die Nutzer zugehen oder solche Dinge abfragen?

Man sollte irgendwie zusammenfinden. Ich weiß, die zuständigen Kollegen, die damit beschäftigt sind, Bestände für die Digitalisierung auszuwählen, suchen solche Kontakte, aber so richtig, dass man daraus in den letzten Jahren Input ziehen konnte, kam nichts. Und das würde ich gerne intensiver sehen.

[Herr Siegmann] Ich sehe hier auch eine nach wie vor bestehende Kluft zwischen einem traditionellen lesenden Bibliotheksnutzer und den Möglichkeiten, die die Digitalisate im Rahmen von Digital Humanities bieten, was aber immer noch zum einen IT-Kenntnisse und zum andern eine gewisse Verankerung in einer bestimmten Szene voraussetzt. Ich glaube, hier sind noch Brücken notwendig, die von beiden Seiten gebaut werden müssen. Bibliotheken sollten als Einrichtungen möglichst niedrigschwellige Tools anbieten, die man für die wissenschaftliche Auswertung von Quellen als Hilfsmittel verwenden kann. Die Digital Humanities sollten sich mit ihren Angeboten stärker auf die traditionelleren Arbeitsweisen von Wissenschaft hinzubewegen. Hier besteht das Bedürfnis nach Zusammenwachsen und das ist nicht nur dadurch erfüllt, dass man einfach Images anbietet, sondern da müssten wir als Einrichtung auch noch weiter gehen, was perspektivisch sicher kommt, aber aktuell auch nicht an erster Stelle steht.

I-7.6. Abschluss: Welche Möglichkeiten (Digitalisierungstempo, Erschließungstiefe, Funktionen, ...) hätte Ihre Einrichtung bei unbegrenzten Mitteln?

[Herr Siegmann] Im Bezug auf die Retrodigitalisierung würde man natürlich viel schneller, viel mehr und viel tiefer arbeiten können. Gerade wenn man jüngere Nutzergruppen durch das Digitalisierungszentrum führt, gibt es gelegentlich Verwunderung, dass noch nicht alle Bestände dieser Bibliothek digitalisiert sind und dass man noch nicht alles, was es gibt, online finden kann. Das wäre natürlich ein wünschenswertes Ziel, dass möglichst alles, was hier vorliegt und rechtfrei ist, auch digital verfügbar ist, so unrealistisch das auch ist. Aber bei unbegrenzten Mitteln käme man dem natürlich wesentlich schneller näher.

[Frau Müller] Uns ist aber schon klar, was für ein gigantisches Unternehmen das wäre. Es ist ja eben nicht nur die Bildproduktion, sondern wie vorhin schon gesagt, auch die Fachabteilungen, die dann katalogisieren, Strukturdaten erfassen, etc. Also es wäre absolut gigantisch, wenn wir dann auch in diesem Stil weiterarbeiten wollten, was wir ja wollen.

Es wäre schon schön, wenn noch viel mehr unserer prominenter Bestände, die analog wie digital bekannt sind, vielleicht noch bekannter werden oder einen größeren Stellenwert in der Öffentlichkeit gewinnen könnten, das wäre natürlich phantastisch. Ich denke z.B. an unsere Zauberflöte oder große Bachhandschriften oder ähnliches. Das wäre natürlich wunderbar und für jeden zugänglich und damit eben noch präsenter als vorher, wo vielleicht ein Wissenschaftler wusste, dass sich diese Unikate in der Staatsbibliothek befinden.

[Herr Siegmann] Was mir noch einfällt, wäre die Vermittlung der Inhalte. Was wir ja im Moment, wenn man mal ganz ehrlich ist, eigentlich machen ist, ein Buch auf einen

Scanner zu legen und dann haben wir von diesem Buch eine digitale Form. In der digitalen Sammlung sind die Bücher auch nur unwesentlich stärker strukturiert als zuvor im Magazin oder im Katalog. Dieser Wunsch ist mir ganz oft schon an unterschiedlichen Stellen begegnet, z.B. im Europeana-Projekt zum Ersten Weltkrieg, wo wir zumindest versucht haben, bestimmte Materialarten separat zu kennzeichnen und ganz grob Themen zu filtern. Das waren dann nur Oberbegriffe wie z.B. „Propagandamaterial“. So schwierig das auch ist, aber das war schon sehr willkommen. Trotzdem ist die dort entstandene Klassifikation nur eine sehr grobe, oberflächliche Klassifizierung. Auf dem E.T.A. Hoffmann-Portal gibt es weiterführende Informationen zum Leben von E.T.A. Hoffmann und zu Zeitgenossen und das wird von den Nutzern extrem positiv angenommen, dass man einen zentralen Anlaufpunkt hat, der vermittelt, dass alles, was zu dieser Person relevant ist, hier zu finden ist. Egal, ob es auch auf dem Server physisch liegt oder mit einem Link oder einer Quellenangabe verwiesen wird. Ich glaube, hier liegt eine Aufgabe, die für Bibliotheken hoffentlich immer relevanter werden wird. Genau diese Bezüge herzustellen, Einstiegspunkte für bestimmte Themen und Sachverhalte zu liefern und nicht nur Materialien zur Verfügung zu stellen. Und das könnte man natürlich zu jedem unserer Projekte wunderbar machen.

9.9.3. Interview Universitätsbibliothek der Humboldt-Universität zu Berlin.

Historische Sammlungen und Digitalisierung, Universitätsbibliothek der Humboldt-Universität zu Berlin - Yong-Mi Rauch

12.06.2019

If-2. Digitalisierung.

If-2.1. Werden bestimmte externe Standards zur Digitalisierung in Ihrer Einrichtung angewandt (z.B. die DFG-Praxisregeln zur Digitalisierung) oder folgt man hausinternen Richtlinien?

Die DFG-Praxisregeln bilden den Mindeststandard. Je nach Objekten gibt es noch höhere Qualitätsstandards, z.B. höhere Auflösungen. Andere Standards werden eher projektbezogen erstellt, je nachdem, was das Material erfordert. Aber überwiegend haben wir ja textbasierte Materialien und die sind nach DFG-Praxisregeln oder höher bearbeitet.

If-2.2. Welche Standards gelten konkret für die Erzeugung von Master-Bilddigitalisaten (z.B. DPI-Anzahl, Farbtiefe, Farbraum, Format, ...).

Mindestens 300 dpi, teilweise höher, Farbraum RGB, im TIFF-Format, teilweise zudem Angebote in Präsentationsformaten.

If-2.3. Wie viele Digitalisate bieten Sie insgesamt an?

Das ist schwer zu sagen. Ungefähr haben wir geschätzt so 7000 Digitalisate.

If-2.4. Wie viele Objekte werden pro Jahr ungefähr neu digitalisiert?

Ungefähr werden pro Jahr 1000 neue Digitalisate publiziert, dies können aber teilweise auch ältere Digitalisate sein, die bisher unbearbeitet gespeichert waren und nun publiziert werden. Der Digitalisierungsprozess wird dadurch abgeschlossen, da das Publizieren ja den letzten Schritt der Digitalisierung darstellt.

If-2.5. Befinden sich Born Digitals in Ihrer Sammlung?

Weniger. Es gibt einige z.B. im Bereich der Porträtsammlung, da einige Professorenporträts digital abgegeben wurden. Wenn man das Archiv hinzuzählt muss man sagen, dass es zwar noch kein richtiges digitales Archiv gibt, man hier aber zunehmend mit der Frage des Umgangs mit digitalen Akten konfrontiert ist. Dieser Prozess hat aber noch keine Routine.

If-2.6. Findet die Digitalisierung im Haus oder durch Dienstleister statt?

- *Wenn im Haus digitalisiert wird: Wie viele Mitarbeiter sind mit der Digitalisierung beschäftigt?*

Sowohl als auch. Für die Digitalisierung im Haus sind drei Personenäquivalente vorhanden, verteilt auf verschiedene Mitarbeiter. Wenn man die EDV hinzuzählt vielleicht viereinhalb Personenäquivalente, aber davon ist keine hauptamtlich nur für die Digitalisierung zuständig.

If-2.7. Gibt es Maßnahmen, um Doppeldigitalisierungen von Objekten zu vermeiden?

- Finden Absprachen mit anderen Gedächtniseinrichtungen statt, um hier bereits digital vorliegende Werke nicht erneut zu digitalisieren?

In der Regel wird das konsequent überprüft. Die Frage, ob das Objekt bereits woanders digital vorhanden ist, ist Teil des Workflows und des Fragenkataloges zur Digitalisierung. Insofern ist das schon eine wirksame Variante, um Doppeldigitalisierungen zu vermeiden. Man muss aber sagen, dass wir die historischen Bücher nicht mehr unter dem Fokus sehen, dass Bücher als Objekte vielfach vorhanden sind, sondern Bücher verschiedene Einbände und verschiedene Gebrauchsspuren, also unterschiedliche Objektgeschichten haben. Insofern digitalisieren wir doch Bücher, die bereits in anderen Beständen vorhanden sind. Zum einen auf Grund der oft mangelhaften Qualität von Massendigitalisierungen, zum andern, da Bücher in bestimmte Provenienzzusammenhänge gehören.

If-3. Dokumentation, Metadaten, Erschließungsstandards.

If-3.1. Welche Metadaten werden bei der Digitalisierung erfasst?

Es gibt ein Grundmetadaten-set, in dem die Daten bibliographisch sowie physisch erfasst werden. Also bibliographische Metadaten wie Titel, Autor, Zeit, Verlag sowie typische Strukturdaten wie Einband, Vorblatt, Titelblatt, Paginierung, Kapitel etc. In der Regel hat dies eine gewisse Tiefe, so dass auch Abbildungseinheiten und Kapiteleinheiten als Metadaten erfasst werden. Wenn wir aber z.B. Zeitschriften digitalisieren, so können wir in der Regel keine tiefe Erschließung auf Titelebene durchführen. Bei umfangreichen Zeitschriften haben wir z.B. eine übergreifende Metadatenerschließung auf Jahrgangs- oder Heftebene. Wir versuchen also Aufwand und strukturierte Navigation gegeneinander abzuwägen. Zusätzlich erhalten die Dokumente eine Sacherschließung, werden also Wissensgebieten zugeordnet, die an der RVK orientiert sind, diese aber nicht komplett abbilden. Die RVK-Stellen werden dabei in Wissensgebiete, also fachliche Begriffe, übersetzt.

Gleichzeitig werden die Bücher möglichst sammlungsbezogen digitalisiert. Es gibt zwar eine Menge sog. Einzeldigitalisate, die z.B. über Nutzerbestellungen kommen oder aus Bestandserhaltungsgründen digitalisiert werden müssen, die dann passenden Sammlungen zugeordnet werden. Aber wenn wir übergreifend digitalisieren, so versuchen wir in Zusammenhängen zu digitalisieren, nicht ein Einzelwerk nach dem andern, da sonst sehr disparate Bestände entstehen.

If-3.2. Stehen alle erfassten Metadaten dem Nutzer frei zur Verfügung?

- Wenn nein: Welche nicht?

Es gibt bestimmt administrative Metadaten, welche eventuell nicht frei zugänglich sind, da sie in der Regel auch nicht von Interesse sind. Diese sind dann aber auch nicht für die Bibliothekare sichtbar. Es gibt also keine internen Notizen, die für den Nutzer nicht zu sehen wären.

Zwar haben spezielle Sammlungen wie Exlibris abweichende sammlungsspezifische Metadaten. Hier sind z.B. Motive erfasst. Auch Provenienzzspuren werden oft erfasst. Das Kernmetadaten-set beschreibt dabei die Objekte und ist daher aber auch zugänglich.

If-3.3. Nach welchen Metadatenschemata werden die Digitalisate beschrieben?

Die Erfassung erfolgt im METS/MODS-Format.

If-3.4. Nach welchen Erschließungsstandards werden die Daten beschrieben? Kommen Normdaten zum Einsatz?

Die bibliographischen Metadaten kommen allein schon durch den bibliographischen Datensatz zum Einsatz, sind also in die GND eingebunden. Auch Identifier werden vergeben. Geographika ordnen wir im Moment nicht zu, das liegt aber daran, dass wir keine Bestände haben, die nach GeoData erfasst werden, auch wenn das in manchen Projekten angedacht ist, z.B. in der Ethnologie, also dem FID Sozial- und Kulturanthropologie. Hier werden dann auch zusätzliche Daten erschlossen, was eine Suchordnung ermöglicht.

In der Erschließung nutzen wir auch den Thesaurus für alte Drucke vom GBV, der auch Gattungsbegriffe erfasst. Dieser wird auch für die Präsentation der Digitalisate eingesetzt.

If-3.5. Wird wenn möglich auch der Volltext erschlossen?

- Wenn ja, in welchem Format wird dieser gespeichert?

Ja, wir erschließen den Volltext. Gespeichert wird dieser im ALTO-XML-Format.

If-4. Veröffentlichung.

If-4.1. Unter welchen Lizenzen sind die Digitalisate zugänglich?

- Welche Lizenz wird favorisiert? Warum?

Die meisten Digitalisate sind unter CC-BY-SA-NC zugänglich. Das ist unsere Grundlizenz. Es gibt einzelne Lizenzen, die strenger sind, da die Digitalisate nicht rechtefrei sind. Hier haben wir für die Publikation Verträge mit Verlagen geschlossen. Und wir haben einige Digitalisate, die noch nicht lizenziert sind, da sie auf speziellen Leseplätzen angeboten werden sollen und unter Umständen andere Lizenzen als die Creative Commons haben, sondern begrenzt zugänglich sind.

Zwischenfrage: Gibt es einen Grund, warum diese Lizenz favorisiert wird?

Es war schon der Wunsch da, eine CC-Lizenz vergeben zu können, aber einige Objekte sollen nicht frei genutzt werden, da diese eine Zitierung erhalten sollen. Zudem gibt es durchaus auch Anfragen von Verlagen für eine kommerzielle Nutzung von den Objekten. In dem Fall möchten wir die kommerzielle Nutzung gerne steuern können. So besteht z.B. für Forschungsprojekte mit kommerzieller Nutzung oder Unterrichtsmaterialien die Möglichkeit, durch Verhandlungen Ausnahmen zu gewähren, es soll aber durch die Lizenz erst einmal eine Kontaktaufnahme erfolgen.

If-5. Interoperabilität und Schnittstellen, Portale.

If-5.1. Können die Metadaten über maschinenlesbare Schnittstellen geharvestet werden?

- Wenn ja, welche Schnittstellen werden angeboten?

Hier gibt es Diskussionen, aber über den Goobi-Server ist es nicht möglich.

If-5.2. Werden die Digitalisate auch in andere Angebote oder Portale (z.B. die Deutsche Digitale Bibliothek/Europeana) eingespielt?

- Wenn ja, in welche Portale spielen Sie die Daten ein? Warum gerade in diese Portale?

- Wenn nein, warum nicht?

Ja, sie werden in die DDB eingespielt. Jedenfalls theoretisch, in der Praxis gibt es noch ein technisches Problem. Aber der Vertrag ist geschlossen, so dass alle freien Daten in die DDB und über diese in die Europeana eingespielt werden sollen.

Die DDB ist einfach das geeignete Portal für Bibliotheken und Archive und bietet sich vom Zuschnitt an. Teilweise ist die Nutzung der DDB auch eine Voraussetzung für die Förderung.

Es gibt auch das ZVDD, mit dem wir die Datenübertragung vorbereitet haben. In der Bedeutung ist dieses zwar etwas hinter die DDB zurückgetreten, aber eigentlich werden die Digitalisate auch hier angeboten.

If-7. Organisatorisches.

If-7.1. Welche Fördereinrichtungen (DFG, BMBF, EU, Land Berlin, ...) unterstützen Ihre Einrichtung?

Im Bezug auf die Digitalisierung derzeit die DFG. Durch den Bund hatten wir im Bereich Digitalisierung noch keine Förderung. Und das Land Berlin hat einen Programmschnitt, der die Universitätsbibliotheken nur in zweiter Linie fördert. Wir hatten zwar Verhandlungen, es kam jedoch nie zu einer konkreten Förderung.

If-7.2. Besteht die Möglichkeit, neue Funktionen in das System der digitalen Angebote zu integrieren?

- Wenn ja, wie oft geschieht dies?

Es findet eine regelmäßige Weiterentwicklung des Systems statt. Es gibt dadurch immer wieder Updates, z.B. von der Präsentation oder den Funktionen, um die Nutzung zu aktualisieren. Es gibt hier jedoch keinen regelmäßigen Turnus. Das letzte Update der gesamten Oberfläche fand Anfang 2018 statt, seit dem gab es aber immer wieder kleinere Updates. Diese laufen in kleineren Schritten und kommen vom Dienstleister her gesteuert.

If-7.3. Gibt es Erhebungen darüber, welche Nutzergruppen Ihre digitalen Angebote (am meisten) nutzen?

Nein, dazu gibt es keine allgemeinen Erhebungen. In einzelnen Bereichen wie den FIDs haben wir zwar Erhebungen, diese sind jedoch nicht auf das gesamte System übertragbar.

I-1. Auswahl.

I-1.1. Wie sieht der Auswahlprozess von Objekten für Digitalisierungsprojekte aus?

- *In wie fern sind Kuratoren/Sammlungsleiter und Konservatoren beteiligt?*
- *Welche Rolle spielen Schäden an den Objekten, z.B. besonders für oder gegen eine Digitalisierung?*
- *Fällt die Entscheidung vor allem durch Mittelgeber oder nach eigenen Bedürfnissen?*
- *In wie Weit werden die Nutzerbedürfnisse berücksichtigt bzw. die Nutzer eingebunden?*

Die Digitalisierung hier im Haus ist ja in der Abteilung Historische Sammlung angesiedelt. Insofern werden die Digitalisierungsprojekte innerhalb des fachlichen Bereichs direkt gesteuert. Es gibt innerhalb des Hauses keine anderen Stellen, die Digitalisierungsprojekte koordiniert. Natürlich gibt es noch die Fachbereiche wie die FID, die die Digitalisierung nach fachlichen Kriterien steuern und es gibt sammlungsbezogene Digitalisierungsprojekte, aktuell z.B. Digitalisierung von die Universitätsbibliothek betreffenden Archivalien. Hinzu kommen Anstöße aus der Universität, z.B. wissenschaftliche Projekte, die an bestimmten Beständen interessiert sind und in Einzelfällen weiteren Einrichtungen der Universität, die jedoch nicht Teil der Bibliothek sind, z.B. Archivalien aus anderen Universitätssammlungen. Die werden bei uns auch bearbeitet, aber nicht in großem Umfang, da wir dafür keine Kapazitäten haben.

Da wir auf Fördermittel zur Digitalisierung angewiesen sind, spielen Geldgeber doch eine entscheidende Rolle. Wir reagieren bei der Antragsstellung auf die Ausschreibungen und Förderrichtlinien und stellen dementsprechende Anträge.

Es gibt also eine laufende Grunddigitalisierung und zudem die Projekte, welche sich nicht nach dem richten, was wir gerne digitalisieren würden, sondern was gefördert wird und so wieder Ressourcen bindet.

Unsere Zielgruppe sind vor allem wissenschaftliche Projekte oder Wissenschaftler, wir können aber nicht alles digitalisieren, was sich die Wissenschaftler wünschen, da das auch immer eine Frage des Volumens ist und ob die jeweiligen Nutzer finanzielle Unterstützung für ihre Wünsche mitbringen können. Wenn z.B. umfangreiche Zeitungsbestände digitalisiert werden sollen, dann müssen die Mittel dazu schon mitgebracht werden. Aber natürlich versuchen wir, Fragen nach solchen Beständen zu unterstützen und bei kleineren Beständen die Digitalisierung im Rahmen unserer Kapazitäten zu realisieren.

Schäden an Objekten spielen in so fern eine Rolle, dass wir Bestände, die beim Digitalisieren Schaden nehmen, nur in sehr wenigen Einzelfällen abwägend zulassen. Z.B. dann, wenn die Objekte so stark geschädigt sind, dass sie nicht mehr genutzt werden können und es auch keine Chancen gibt, diese konservatorisch zu erhalten. Da diese Objekte nur noch einmal digitalisiert werden können, aber auch sonst nicht mehr benutzt werden dürfen, bewahrt man sie damit digital für die wissenschaftliche Nutzung und kann das Original wegschließen. Dies ist aber wie gesagt ein großer Einzelfall. Wenn wir sonst Schäden haben, die durch die Digitalisierung verstärkt werden würden und es andere Möglichkeiten gibt, das Original zu bewahren oder es nicht nötig ist, das Original zu digitalisieren, da es bereits wo anders vorliegt, dann würden wir von der Digitalisierung absehen. Dies betrifft vor allem historische Drucke, die wegen ihres Öffnungswinkels oder ihres Einbandes nicht digitalisiert werden können. Dann würde der Schutz des Originals

vor der Digitalisierung stehen. Alle anderen Formen wie das Aufbinden von Bänden würden einen sehr großen Materialeingriff bedeuten und dies wäre nur gerechtfertigt, wenn es gravierende Gründe gibt, die den physischen Erhalt im aktuellen Zustand überwiegen würden. Das ist aber sehr theoretisch, bei uns ist der Fall bisher noch nicht eingetreten.

Es gibt einen Geschäftsgang zur Digitalisierung aus Bestandsschutzgründen. Diese Bestände werden so gesichert und dann der Benutzung entzogen. Dieser Geschäftsgang wird zwar genutzt, aber nicht sehr stark, da unsere am meisten beschädigten Bände meistens in den Zeitraum fallen, der rechtlich noch problematisch ist, sie sind also noch nicht rechtfrei und wir könnten sie nur beschränkt bereitstellen.

I-1.2. Wie sieht der Umgang mit Dubletten aus: Werden beide Exemplare, z.B. eines Buches, digitalisiert, wird ein beispielhaftes Objekt aus beiden Dubletten erstellt oder wird sich für eines entschieden?

Ich würde auf jeden Fall eine Mischung aus mehreren Exemplaren ablehnen, es sei denn, es gibt gute Gründe dafür. Würde man z.B. eine komplette Zeitschrift wiederherstellen wollen, dann könnte man so vielleicht ein komplettes Exemplar mit allen Seiten herstellen, wenn der komplette Zeitschriftenverlauf in keiner Bibliothek in Deutschland mehr vorliegt. So einen Fall hatten wir und hier haben wir wirklich verschiedene Exemplare aus Bibliotheken gemischt und eine vollständige Zeitschrift zusammengestellt. Dies ist aber kein regulärer Workflow.

Hinzu kommt, dass es gerade in den historischen Beständen keine echten Dubletten, sondern nur mehrere Exemplare einer Ausgabe gibt. Man entscheidet sich in der Regel für ein Exemplar, je nach Bestand kann es aber auch sein, dass man beide Exemplare digitalisiert. Wir haben z.B. Bücher, die sind zweimal in der Bibliothek der Gebrüder Grimm, je mit Widmung einmal für Jacob und einmal für Wilhelm Grimm. Wenn ein Nutzer nun nach einer Digitalisierung fragt, so würde man sich alle Exemplare anschauen und dann für eines entscheiden, wobei man natürlich versucht, ein besonders schönes Exemplar bzw. eines, was von den zusätzlichen Merkmalen besonders interessant ist für die Digitalisierung auszuwählen. In dem Fall haben wir uns dann für das Exemplar von Jacob Grimm entschieden, da dieses zusätzlich auch handschriftliche Annotationen von ihm enthält. Das wäre ein Entscheidungsprozess, aber wir würden nicht systematisch alle Exemplare eines Werkes digitalisieren, da wir dann wirklich digitale Dubletten hätten mit dem gleichen Text, bei denen sich vielleicht nur der Einband unterscheidet. Dies wäre nicht ökonomisch.

I-2. Digitalisierung.

I-2.1. Erfolgt die Digitalisierung immer vom Original oder wenn vorhanden auch vom Film/Microfiche?

Wir stellen eigentlich keine Digitalisate vom Microfiche her und haben selber dafür hier auch keine Technik, man müsste dies also an einen Dienstleister geben. In der Regel versuchen wir eine Einzeldigitalisierung von wertvollen oder unikalen Beständen durchzuführen in einer Qualität, von der wir hoffen, dass man sie nicht mehrfach

durchführen muss. Also eine hochwertige Digitalisierung, die auch Teile des Einbandes, Leerseiten, buchspezifische Merkmale wie Spiegel usw. beinhaltet, in der Hoffnung, dass diese Digitalisate auch viele Fragen über den Text hinaus lösen können. Also dass man nicht nur sehen kann, was im Text steht, sondern auch, wie das Exemplar aussieht, ob es Provenienzmerkmale hat usw. Deswegen wäre eine Digitalisierung vom Film eh nur zweite Wahl, da viele dieser Merkmale hier nicht erkennbar sind und die Wiedergabe in der Regel nicht überzeugend ist. Es wäre höchstens eine Option wenn man Publikationen hat, die sonst ein kompletter Verlust sind. Diesen Fall hatten wir aber noch nicht.

I-2.2. Als Qualitätsstandard werden oft die DFG-Praxisregeln zur Digitalisierung genannt. Wie relevant sind diese für die Digitalisierung in ihrer Einrichtung?

Die Frage hatten wir ja schon geklärt.

I-2.3. Welche Maßnahmen zur Qualitätssicherung werden getroffen?

Die Digitalisate durchlaufen alle eine doppelte Qualitätskontrolle. Sie werden direkt nach dem Scanvorgang durch eine andere Person als den Scanoperator auf Vollständigkeit und Qualität kontrolliert. Dann erfolgt in der Metadatenerfassung auch etwas wie eine automatische Kontrolle, ob die Paginierung richtig sein kann. Hier werden die Digitalisate auch nochmal angeschaut, es wird hier nochmal die Qualität kontrolliert. Die Publikation wird dabei immer durch bibliothekarisches Fachpersonal und nicht durch Hilfskräfte erstellt und das ist auch ein wichtiger Punkt bei der Qualitätssicherung.

I-2.4. Werden Daten im Nachhinein noch verändert?

- Wenn ja, findet eine Erfassung der Veränderungshistorie statt?

Wie gesagt, für die Präsentation wird das Format verändert, da die TIFFs nicht als Präsentationsformat genutzt werden können. Die Daten werden in der Regel aber nicht bearbeitet. Es gibt dazu Vorgaben für die Digitalisierung, die allen Mitarbeitern als Dokumentation zur Verfügung steht. Dort werden auch Qualitätsstandards für die Digitalisierung festgelegt, so dass immer die komplette Seite mit einem kleinen Teil des Falz und Seitenrand gescannt wird, so dass man auch sehen kann, wie die Maße der Seite sind. Die Seiten müssen während des Digitalisierungsprozesses gut gescannt werden, da sie nachher nicht mehr bearbeitet werden. Auch die Beschaffenheit des Papiers wird nicht durch Kontrastabgleiche etc. verbessert, sondern wir versuchen, die Wiedergabe so nah wie möglich am Original erscheinen zu lassen. Dementsprechend findet eigentlich wenig eine Nachbearbeitung statt.

I-3. Dokumentation, Metadaten, Erschließungsstandards.

I-3.1. Welche Informationen zum Kontext der Digitalisierung (Sammlungsbereich, Provenienz, Digitalisierungsprojekt, ...) werden dokumentiert und verfügbar gemacht?

Es gibt einzelne Projektbeschreibungen, die auch auf der Projekthomepage publiziert werden. Dieser Bereich ist bei uns in der Entwicklung. Zu den Sammlungen gibt es Beschreibungen mit dem Hinweis, was bereits digital vorhanden ist, auch mit Verlinkung zu den Digitalisaten. Geplant ist zudem eine Seite, wo die digitalen Sammlungen

beschrieben werden. Dies ist jedoch noch nicht realisiert, es sei denn, es ist Teil einer anderen Sammlungsbeschreibung. Aber das Ziel ist eigentlich, unter den Digitalisaten selbst nochmal die Sammlungen bzw. Digitalisierungsprojekte zu beschreiben, um dort noch mehr Zusatzinformationen zu bieten. Und diese werden dann auch vollständig verfügbar sein.

Intern haben wir natürlich bei Projektanträgen die Dokumentation und die Arbeitsunterlagen, welche teilweise sehr umfangreich sind, wie z.B. Listen der bearbeiteten Objekte mit Zustandsbeschreibungen.

Der Kontext der Digitalisierung und Hinweise zu Provenienzen etc. ist bei uns durch die Zuordnung zu bestimmten Sammlungen in der Präsentation realisiert.

I-3.2. Durch die Digitalisierung gehen einige Informationen wie z.B. die Materialität der Quelle in der Nutzung verloren. Können umfangreiche Metadaten hier einen gewissen Ausgleich schaffen?

Ich glaube, dass Digitalisate die Materialität nur bis zu einem bestimmten Grad ersetzen können. Durch die hohe Qualität inzwischen jedoch zu einem sehr hohen Grad, da man durch die Vergrößerungsfunktion usw. sehr nah an die Digitalisate herankommt. Wichtig dazu ist, was man ja eigentlich immer dabei hat, so etwas wie ein Farbkeil, der auch die Größenordnungen angibt. Was mich bei den meisten Digitalisaten eigentlich stört, ist, dass man die Größenverhältnisse nicht mehr hat. Gerade bei einem Buch ist es ja interessant, ob es von der Materialität her ein kleiner Duodez Band oder ein Großfolio Band ist. Und diese Proportionen gehen in der Digitalisierung verloren. Manchmal ja auch zu Gunsten der Objekte, die dann auf einem beleuchteten Bildschirm schön groß und farbig wirken, eigentlich aber klein und matt sind. Dafür hat man mit den Digitalisaten natürlich auch tolle Arbeitsmöglichkeiten wie die Vergrößerungsfunktion. Wir versuchen diesem Proportionsverlust entgegen zu wirken, indem wir alles vom vorderen bis zum hinteren Buchdeckel digitalisieren, teilweise auch mit Schnitt und Rücken, wenn es dort etwas besonderes zu sehen gibt. Dies ist jedoch ein weicher Faktor, wir können bei unserem Bestand nicht den Schnitt aller Bücher von allen Seiten scannen, solange dort nicht weitere interessante Informationen zu entdecken sind.

Weitere Metadaten zur Objektbeschreibung geben wir hier nicht an, außer bibliothekarische Standardmetadaten, wie z.B. fehlende Seiten. Aber das ist ja keine echte Beschreibung der Materialität.

Zwischenfrage: Und wenn man an Metadaten denkt, die man bisher noch nicht erfasst?

Das kommt natürlich immer darauf an, welche Materialien wir digitalisieren, aber die wichtigen Metadaten werden hier eh erfasst. Z.B. die Technikbestimmung in der Porträtsammlung, aber diese findet unabhängig von der Digitalisierung statt. Ansonsten fällt mir auch kein Fall ein, wo zusätzliche Metadaten erfasst werden, um diesen Verlust von Materialität durch die Digitalisierung auszugleichen.

I-4. Veröffentlichung.

I-4.1. Werden alle digitalisierten Objekte (soweit rechtlich möglich) auch in Ihren digitalen Angeboten zur Verfügung gestellt? Sind hier alle angebotenen Digitalisate frei zugänglich?

Ja, werden sie mit der Einschränkung der rechtlichen Möglichkeiten. Bei der Porträtsammlung werden Thumbnails und Digitalisate in abgeschwächter Qualität angeboten. Aktuell sind wir zudem dabei, einen elektronischen Leseplatz für rechtlich geschützte Inhalte bereitzustellen.

I-4.2. Werden die Digitalisate in der höchstmöglichen Qualität (Master) in der digitalen Sammlung Ihrer Einrichtung angeboten?

- Wenn nein, warum nicht, worin unterscheiden sich die angebotenen Digitalisate vom Master?

- Wenn nein, was ist nötig, um die Master-Dateien zu erhalten?

Die Digitalisate haben im Internet eine geringere Größe und werden nicht als TIFF angeboten. Sie sind aber für die meisten Gebrauchsanwendungen immer noch gut zu nutzen, bis hin zur Präsentation im Netz oder in einem Vortrag. Für den Print würde diese Qualität knapp nicht ausreichen. Für Daten in hoher Qualität müssen die Digitalisate nach der Gebührenordnung bestellt werden, so dass eine Nutzungsvereinbarung getroffen werden kann, wobei es hier vor allem um die korrekte Zitierweise geht.

I-4.3. Welche Rolle spielen Zertifikate für Forschungsdatenrepositorien (z.B. Data Seal of Approval oder das DINI-Zertifikat) bei Ihren digitalen Angeboten?

Unser Digitalisate-Server ist nicht zertifiziert.

I-5. Interoperabilität und Schnittstellen, Portale.

I-5.1. Findet eine Verlinkung zwischen Objekten (intern wie extern) statt?

Nein, eine Verlinkung findet nicht zwischen den Objekte selber, sondern nur zwischen den Metadaten statt.

I-5.2. Werden die Digitalisate von Ihrer Seite auch auf weniger wissenschaftlichen Plattformen wie Wikipedia angeboten?

- kurze Begründung

Naja, ein paar Bilder werden in der Wikipedia genutzt, diese wurden aber nicht von unserer Seite dort hochgeladen. Wir haben aber tatsächlich Digitalisate auf weniger wissenschaftlichen Plattformen, da wir selber einen Flickr-Account haben und dort ausgewählte Bestände in Alben anbieten. Dies ist ein Mittel zur Öffentlichkeitsarbeit, um ein Schaufenster zu haben und die Bilder einem breiteren Publikum präsentieren zu können, gerade auch bei Dingen, die nicht wissenschaftlich relevant, aber attraktiv sind. Der Account entstand im Zuge einer Überlegung zu virtuellen Ausstellungen. Dabei hatten wir festgestellt, dass viele virtuelle Ausstellungen sehr aufwändig kuratiert werden müssen, vielleicht etwas zu aufwändig, um ein Publikum zu erreichen, was sich nicht durch eine Ausstellung über mehrere Ebenen klicken möchte und wir zudem nicht die Ressourcen hatten, so eine Ausstellung aufwändig mit begleitenden Texten etc. zu

kuratieren. Zudem sind viele dieser Angebote kostenpflichtig und wären nur schwer in unseren Webauftritt einzubinden gewesen. Daher kamen wir dann auf Social Media Angebote und entschieden uns für Flickr, da man sich hier nicht anmelden muss, um die Angebote anzuschauen, was bei den meisten anderen Plattformen anders ist. Diese Anmeldung finde ich für eine öffentliche Präsentation sehr nachteilig und für eine öffentliche Einrichtung nicht angemessen. Zudem muss man sagen, dass Flickr damals noch sehr wenig Werbung gezeigt hat, was seit einem Besitzerwechsel inzwischen leider mehr geworden ist.

I-6. Speicherung, Pflege und Archivierung.

I-6.1. Gibt es spezielle Strategien (z.B. regelmäßige Migration, Konvertieren, ...) zur Langzeitarchivierung der Daten?

Es gibt an der HU selber nur eine Strategie zur Langzeitspeicherung, also eine multiple Speicherung der Daten an verschiedenen Orten Berlins. Zusätzlich steht die Universitätsbibliothek für die Langzeitarchivierung von Daten in Kontakt mit dem Konrad-Zuse-Zentrum Berlin und hat hier den Status eines Pilotpartners für das EWIG-Projekt. Dies ist jedoch noch ein Experiment und keine Routine und es ist durchaus möglich, dass sich die Universitätsbibliothek langfristig einen anderen Anbieter zur Langzeitarchivierung der Daten sucht. Idee ist es, ein kooperatives übergreifendes Angebot zu nutzen statt eine eigene Strategie aufzubauen, da dies nicht der Leistungsbereich der UB ist.

I-6.2. Wie sieht die Backup-Strategie für die Digitalisate aus?

- In welchen Abständen werden Sicherungskopien durchgeführt?
- Werden redundante Sicherungskopien geführt?
- Werden sie in anderen Formaten als im laufenden System gesichert?

Das ist Aufgabe der EDV und nicht mein Arbeitsbereich. Ich weiß aber, dass die Redundanzsicherungskopie vierfach abläuft, aber nicht in anderen Formaten gesichert wird.

I-6.3. Lohnt sich eine langfristige Archivierung von Digitalisaten (im Bezug auf Aufwand, Finanzen, Verwaltung, ...) oder sollten diese bei Bedarf (beschädigte Daten, veraltete Formate, ...) einfach neu vom Original aufgenommen werden?

Ich weiß, dass Archive hier eine andere Strategie verfolgt haben als Bibliotheken. Hier wird eher auf Filme zur Langzeitarchivierung gesetzt, da man diese als analoge Medien einfach in Salzbergwerken etc. einlagern und ohne weitere Hilfsmittel noch lesen kann. Diese Strategie verfolgen Bibliotheken ja nicht mehr. Eine langfristige Archivierung der Digitalisate ist daher lohnend, besonders bei Digitalisaten, die mit einem so hohen Aufwand erstellt werden wie bei uns üblich. Man möchte ja, dass die Digitalisate erhalten bleiben und in möglichst vielen Fragen das Original ersetzen können. Zudem wollen wir dem Original ja nicht eine mehrfache Digitalisierung zumuten müssen, so unsere Hoffnung. Bei Bedarf neue Aufnahmen sollte man möglichst vermeiden, auch wenn es bei beschädigten Daten vielleicht schwierig zu vermeiden ist. Bei den Dateiformaten setzen wir auf Standardformate mit der Hoffnung, dass diese eine gewisse Haltbarkeit besitzen

und sich möglichst wenig verändern bzw. wenn doch, dass dann das Problem einer Massenmigration vorliegt, man also für wenige Formate eine Migration für alle Einrichtungen entwickeln kann. Das ist ja die Strategie, die Bibliotheken in diesem Bereich verfolgen.

Auf keinen Fall sollte man alle zehn Jahr versuchen, die Objekte neu zu digitalisieren. Dies ist einfach ein Verbrennen von personellen und finanziellen Ressourcen sowie der Originale.

I-6.4. Sollten bei einer qualitativ besseren Neuaufnahme diese Digitalisate an Stelle der alten Daten treten oder sollte hier ein neuer Datensatz angelegt und mit dem alten verlinkt werden?

Wir haben natürlich manchmal die Fälle, dass wir mehrfache Digitalisate haben, die z.B. auf mehreren Plattformen sind, z.B. mehrere Angebote auf dem edoc-Server und dem Goobi-Server. Die Frage ist für uns aber nicht wirklich zutreffend, da wir eine Einrichtung sind, die erst sehr spät mit dem Digitalisieren begonnen hat.

Wir haben natürlich auch Altdigitalisate, die damals nicht publiziert wurden, da sie auch nicht nach den (heutigen) Workflowstandards erstellt wurden, also z.B. nur digitalisiert, im Katalog aber keine Metadaten angelegt wurden. Hier muss abgewogen werden, ob das Objekt neu aufgenommen werden muss oder ob man mit den alten Digitalisaten arbeiten kann.

Den Fall, dass das Objekt bereits neu digitalisiert wurde und dennoch Altdigitalisate vorliegen, hatten wir noch nicht. Aber auch hier würde man entscheiden, auch unter der Frage, ob verschiedene Stufen der Digitalisierung für die Wissenschaft relevant sein können.

Wir haben aber eher den Fall, dass wir unpublizierte Altdigitalisate haben und darüber nachdenken, ob hier ein neues Digitalisat nötig ist oder nicht. Hier würde man vom Objekt her entscheiden. Und für den Fall, dass das Objekt neu digitalisiert wird, würden wir wahrscheinlich das neue Digitalisat nutzen, sonst müssten wir das Objekt ja nicht neu digitalisieren, wenn das alte reicht. Besonders, da der volle Digitalisierungsvorgang sehr aufwändig ist.

I-6.5. Werden nicht mehr genutzte oder veraltete Digitalisate (z.B. veraltetes Dateiformat) aus dem System gelöscht?

- Wenn ja, unter welchen Bedingungen genau?

Veraltete Digitalisate im Sinne veralteter Formate haben wir nicht, da sich unsere genutzten Formate im Wesentlichen nicht geändert haben. Im Sinne von qualitativ veralteten Digitalisaten würden wir diese wahrscheinlich löschen, da sie nicht genutzt werden und wir sie nicht gebrauchen können, sie also nur Speicherplatz blockieren. Natürlich würden wir vorher prüfen, ob das Objekt neu digitalisiert werden kann.

Wir würden zudem keine Daten verwalten, die nicht Bestand der HU sind, z.B. wenn es sich um Digitalisate oder Verfilmungen handelt, die irgendwann mal im Zuge eines Forschungsprojektes ans Haus gekommen sind. Dann würde man schauen, ob die Objekte woanders noch liegen. Also man würde keine Digitalisate löschen, deren Originale vernichtet sind, aber in der Regel sichern wir keinen Bestand, der nicht Bestand

der HU ist. Wir machen Retrodigitalisierung unseres Bestandes und nicht für andere Einrichtungen.

I-7. Organisatorisches.

I-7.1. Liegt ein Datenmanagementplan oder ähnliches zum Umgang mit den Digitalisaten vor?

- Wenn ja, welche Aspekte des Datenmanagements werden beschrieben?
- Wenn ja, wer ist für die Umsetzung verantwortlich?

Das wurde von uns so noch nie abgefordert. Gegenüber Drittmittelgebern argumentieren wir hier immer damit, dass wir eine nachhaltige und fest bestehende Einrichtung sind. Das heißt, dass es uns noch nächstes Jahr und auch noch in zehn Jahren geben wird. Wir haben ein Konzept für die Langzeitarchivierung, verwenden Standardformate und Standardvorgänge und richten uns nach Standardworkflows, sowohl für die Software wie für die Datenspeicherung, die auch andere Bibliotheken nutzen. Das wäre der Aspekt unseres Datenmanagements, auch wenn es nicht so verankert ist.

I-7.2. Wer ist Ihre Zielgruppe (Fachcommunity, interessierte Öffentlichkeit, andere Einrichtungen, ...)?

Unsere Hauptzielgruppe ist natürlich die Wissenschaft, besonders unsere Universität. Ein gewisser Teil richtet sich auch an unser gesamtes Bestandskonzept und das Bestandsmanagement. Die interessierte Öffentlichkeit ist nur am Rande unsere Zielgruppe. Wir würden also keine Bestände allein dafür digitalisieren, dass sie in öffentlichen Netzwerken etc. genutzt wird. Unser Ziel ist also ganz klar die Digitalisierung für die wissenschaftliche Nutzung.

Zwischenfrage: Wissenschaft schließt hierbei auch Studenten mit ein?

Ja natürlich. Ich sehe das Studium als Vorbereitung bzw. Ausbildung für die Wissenschaft an. Man kann daher auch sagen Studium und Lehre bzw. Forschung und Lehre. Wobei man ja auch sagen muss, dass die historischen Quellen nur für bestimmte Fachrichtungen relevant sind. Hier ist eher die Frage, in wie weit wir es schaffen, mit diesen Gruppen zusammenzuarbeiten und ob diese sich z.B. auch aktiv an uns zur Digitalisierung wenden. Wir haben manchmal das Problem, dass uns einige Wissenschaftler auffordert, spezielle Bestände zu digitalisieren. Zwar haben wir einen Service zur Auftragsdigitalisierung, dieser ist jedoch nicht kostenfrei, denn wir können nicht auf Zuruf irgendein Werk digitalisieren, da es gerade gebraucht wird. Wir haben dazu nicht die Ausstattung und müssen uns im Sinne der Bestandskonzepte auch fragen, ob wir z.B. aus einer Zeitschriftenreihe nur einen Einzelband digitalisieren wollen, nur weil dieser gerade gebraucht wird. Das ist für uns unbefriedigend. Das ist daher nur dann möglich, wenn Wissenschaftler dazu auch finanzielle Mittel mitbringen. Unser Fokus ist daher wirklich die eigene Einrichtung, da wir ja auch nur für diese ausgestattet sind und leider nicht, um z.B. Forscher deutschlandweit zu versorgen.

I-7.3. Werden bekannte Nutzergruppen bei Entscheidungen (z.B. was als nächstes digitalisiert wird) mit einbezogen?

- Wenn ja, wie findet die Einbeziehung statt?

Das kann man so eigentlich nicht sagen. Also wir fragen nicht herum, was als nächstes digitalisiert werden soll. Wir haben eher die Überlegungen, was wir für Bestände haben, die sonst nirgends vorhanden und wissenschaftlich relevant sind. Diese werden dann nach und nach bearbeitet. Wenn wir natürlich Bestände haben, wo wir eine spezifische Nutzergruppe haben und klar ist, dass diese daran interessiert ist, und wir nur die Kapazitäten für gewisse organisatorische Abschnitte haben, dann würden wir diese Nutzergruppe natürlich mit einbeziehen und fragen, was aus dem Bestand prioritär digitalisiert werden soll. Die Gelehrtenbibliotheken sind ein solcher Fall. Hier wurden die Nutzer direkt angefragt und um Einschätzung gebeten, was digitalisiert werden sollte. Dies trifft wie gesagt aber auf Einzelbereiche zu.

I-7.4. Sollten Einrichtungen die Forschungsdaten, die aus ihren Digitalisaten hervorgehen (z.B. Korpora, Textauswertungen, ...), in Form eines Repositoriums sichern und verfügbar machen?

Ich halte das für eine gute Idee und auch bei Diskussionen oder Anträgen zeigt sich, dass versierte Wissenschaftler oft nicht daran denken, die Bibliothek mit einzubeziehen, sondern lieber mit eigenen Instrumenten arbeiten, bei denen man sich aber fragt, wie da die Nachhaltigkeit, nicht nur die Sicherung, ist und wie diese Daten auch mit Daten einer anderen Community vernetzt werden können. Dies ist ein großes Problem, da unsere UB das mit ihrer Infrastruktur eigentlich nicht leisten kann. Wenn ein Forscher kommt und Daten abgeben möchte, dann müssten diese mit unserer Software verknüpft werden. Dazu bräuchten wir Entwickler oder finanzielle Mittel, oft sind die Daten auch nicht ohne weiteres kompatibel. Das ist sehr problematisch, es gibt hier aber keine Handhabung.

Ein anderer Punkt ist, dass die UB auch nicht dazu ausgestattet ist, dass Einrichtungen oder Wissenschaftler mit Projekten an die UB herantreten und fragen, ob diese sie dabei unterstützen kann, also Support leisten, Speicherplatz bieten, eine Software besorgen, betreuen und entwickeln sowie das Projekt langzeitarchivieren, also ca. auf 30 Jahre sichern, kann. Dazu müsste sie einen Auftrag von Seiten der Universität bekommen, um solche Forschungsprojekte, z.B. im Bereich Digital Humanities, zu betreuen, neben der Beratung ein Konzept für solche Forschungsdaten vorzulegen und anzugeben, wie man diese Daten einbinden könnte. Man müsste also ein Forschungsdatenkonzept entwickeln, obwohl viele Forschungsdaten sehr unterschiedlich sind und viele Wissenschaftler ungern fertige Lösungen übernehmen, sondern meist etwas anderes, viel besseres und viel ausgefeilteres wollen. Dafür ist die UB leider nicht aufgestellt. Es wäre sehr wünschenswert, wenn man diese Projekte betreuen könnte und hier eine Vernetzung zwischen Grunddigitalisat und z.B. Edition schaffen könnte, aber schon das sehe ich als problematisch an, da z.B. ein Digitalisate-Viewer, der für das Zeigen von Digitalisaten mit Volltext gemacht ist, meist keine Editionsfunction hat. Ich habe auch Zweifel, dass es möglich und sinnvoll ist, so etwas darin zu implementieren, da man dann ganz andere Anforderungen an dieses Repositoryum bzw. diesen Viewer stellen würde. Es ist eher sinnvoll, dann eine andere Plattform anzulegen, von der die Daten geladen und weiterbearbeitet werden können. Und dazu bräuchte man eine Sicherungsfunktion. Das wäre sinnvoll. Ob die UB dies dann betreuen müsste ist eine andere Sache, aber man

würde so die Verbindung zwischen diesen Projekten und den Digitalisaten schaffen. Vielleicht sollte man diese Verbindung auch in mehreren Schritten statt einer einzigen Software realisieren, aber momentan findet es gar nicht statt. Ich würde es sinnvoll finden, wenn es eine stärkere Einbeziehung geben würde und wenn dies Auftrag der UB werden würde. Aber dazu bräuchte man mehr Personal.

I-7.5. Abschluss: Welche Aspekte sollten sich Ihrer Meinung nach bei der Digitalisierung in Gedächtniseinrichtungen verbessern?

Ich glaube, dass sich die Präsentationsschicht verbessern sollte. Diese sind nicht gut genug, gerade für jüngere wissenschaftliche Nutzer, die ganz andere Präsentationen gewohnt sind.

Es wäre auch besser, wenn der Austausch zwischen Nutzern und den Einrichtungen selber stärker wäre. Ich glaube, dass es günstig wäre, wenn die Gedächtniseinrichtungen ihre Nutzer besser in ihre Konzepte einbinden würden. Das Problem dabei ist natürlich, dass es sehr verschiedene Nutzer gibt, so dass es schwierig ist klar zu machen, welche Nutzergruppen man als Einrichtung braucht. Wenn man vorrangig handschriftliche Archivalien nutzt, dann können Funktionen zur Massennutzung interessant sein, es ist aber die Frage, ob diese Nutzer solche Massenverarbeitungen vornehmen oder ob dann nicht eher Forschungsprojekte in anderen Einrichtungen, die z.B. Programme für Handschriften entwickeln, eher die Zielgruppe wären. Es ist oft sehr schwierig, hier die geeigneten Gruppen zu finden, da diese ja auch nicht als Gruppen ansprechbar sind. Daher ist z.B. die Optimierung der Präsentation- oder der Bearbeitungsmöglichkeiten nicht so einfach.

I-7.6. Abschluss: Welche Möglichkeiten (Digitalisierungstempo, Erschließungstiefe, Funktionen, ...) hätte Ihre Einrichtung bei unbegrenzten Mitteln?

Wir könnten viel umfassender digitalisieren, könnten viel besser erschließen und könnten das bisher bestehende Defizit auf der konzeptuellen Ebene ausgleichen. Wir könnten also bessere Konzepte zur Digitalisierung entwickeln und hätten damit oft mehr Möglichkeiten zur Einwerbung von Mitteln, also zur Teilnahme an Projekten. Denn da besteht bisher ein Stau, wir können nicht den Überblick über die gesamte Förderlandschaft behalten und dann dazu passende Anträge generieren.

Weitere Möglichkeiten wären die Mittel zur Ausstattung. Unsere Ausstattung ist bisher im Hardwarebereich nicht ausreichend. Bei unbegrenzten Mitteln könnten wir auch im räumlichen Bereich expandieren. Wir hätten dann auch mehr Möglichkeiten, die Workflows zu beschleunigen und das Angebot zu verbessern, z.B. zur Optimierung des Viewers oder der Nutzungsmöglichkeiten. Ohne die finanziellen Mittel stößt man hier sehr schnell an seine Grenzen.

Für einige Förderrichtlinien benötigt man zudem umfassende Eigenanteile, so dass wir dadurch, dass wir keine festgeschriebenen Projektmitteln haben, eigentlich nicht in der Lage sind, umfassende Projekte zu formulieren.

Über die UB hinaus siehe ich die Möglichkeit, universitätsweit materialübergreifend oder über andere Formate zu digitalisieren, was auch sehr attraktiv wäre, da wir innerhalb der Sammlungen auch Überschneidungen haben. Im Moment sind wir darauf festgelegt,

bestandsbezogen zu arbeiten und es wäre daher schon sehr interessant, sammlungsübergreifend arbeiten zu können, um z.B. die passenden Objektsammlungen einzubinden. Das würde aber eine ganz andere Aufstellung erfordern, weg von dem Konzept der reinen Retrodigitalisierung hin zu einer konzeptuell viel ausgefeilteren Entwicklung von Angeboten. Dann würden wir mit der Sammlung viel enger zusammenarbeiten und könnten die Digitalisierung auch viel besser kuratieren, also z.B. Objekte verlinken. Das können wir im Moment nicht, da es eine andere konzeptuelle Aufstellung bedeuten würde, wenn man die Verlinkung nicht mehr nur zwischen Büchern herstellt, sondern eher eine virtuelle Sammlung zusammenstellen würde. Es wäre dadurch interessanter, andere Objekte heranzuholen. Worüber wir aktuell natürlich auch noch nachdenken ist, ob wir nicht IIF als Schnittstelle anbieten, aber das würde bedeuten, dass wir es dem Nutzer überlassen, solche Verlinkungen anzulegen anstatt solche Verlinkungen selber anzubieten.

9.9.4. Interview Nutzer 1.

Nutzer 1

23.05.2019

Nf-1. System.

Nf-1.1. Ist Ihnen als Sucheinstieg die Suche oder das Browsing/Facettierung wichtiger?

Sowohl als auch. Ich benutze beides und komme auch mit beidem gut klar, weil es immer wieder auch Ergebnistreffer gibt.

Nf-2. Anforderungen.

Nf-2.1. Welche Metadaten eines Digitalisates sollten auf jeden Fall, welche idealerweise erfasst werden?

Auf jeden Fall die bibliographischen Daten. Je detaillierter die Angaben sind, desto eher bin ich geneigt, diese Quelle zu nutzen. Ich ärgere mich gerade bei Bilddateien über einen laxeren Umgang mit den Metadaten. Für mich sind die Metadaten ganz entscheidend, einerseits, um die Glaubwürdigkeit eines Digitalisates unter Beweis zu stellen und andererseits als Grundlage, um überhaupt damit arbeiten zu können.

Nf-2.2. Welche Grundanforderungen stellen Sie an Bilddigitalisate (z.B. Farbigkeit, Vergrößerungsfunktion, Farbraum, Qualität, OCR...)?

Die Grundanforderungen sind sehr wichtig. Die Farbechtheit spielt eine ganz große Rolle. Wenn ich über ein Bild arbeite, würde ich nie über ein Digitalisat schreiben, sondern immer nur über das Original. Ich würde sogar um die Welt reisen, um das Original wenigstens einmal gesehen zu haben und meine Eindrücke zu notieren, um dann weiter mit einem Bilddigitalisat zu arbeiten. Ich würde nie einen Text veröffentlichen, wo ich das Original nicht selbst gesehen habe.

N-1. System.

N-1.1. Nutzen Sie eher Portale (DDB, Europeana, ...) oder direkt die digitalen Angebote von Gedächtniseinrichtungen?

- Können Sie dies kurz begründen?

Ja, ich nutze manchmal Portale, bspw. Europeana. Oft für Unterrichtsmaterial, weniger zur eigenen Forschung. Ich komme seltener über solche Portale zu meinen eigenen Forschungsthemen. Ich bin sehr auf materielle Objekte wie gedruckte Bücher angewiesen.

Aus gesundheitlichen Gründen kann ich einen Großteil der Zeit nicht auf Bildschirme gucken, bin daher in der Zeit begrenzt, die ich am Bildschirm verbringen kann. Auch Texte für meine eigene Forschung lese ich nur digital, wenn es nicht anders geht.

N-1.2. Welche Funktionen in digitalen Angeboten von Gedächtniseinrichtungen nutzen Sie am meisten bzw. sind am wichtigsten für Sie?

Ich habe keine Vorlieben, auch keine Portale von bestimmten Gedächtniseinrichtungen. Solche Suchen gebe ich aus gesundheitlichen Gründen an meine Hilfskräfte, habe also nur einen sehr eingeschränkten, indirekten Zugriff auf solche Angebote.

Zwischenfrage: Wenn Sie Aufgaben abgeben, überlassen Sie Ihren Hilfskräften, wie sie die Informationen suchen.

Genau. Auch wenn ich die Angebote nicht selber nutzen kann, halten mich meine Hilfskräfte jedoch über Neuerungen auf dem Laufenden, so dass ich bedingt auf einem aktuellen Stand bin und weiß, wie sie in diesen Systemen nach Informationen suchen.

N-2. Anforderungen.

N-2.1. Mit was für digitalen Daten arbeiten Sie hauptsächlich?

Mit ganz wenigen, das ist bei mir sehr eingeschränkt. Ich benutze meinen Computer fast immer noch, bis auf wenige Ausnahmen, wie eine bessere Schreibmaschine. Textdokumente, das Verwalten und die Suche in Textdokumenten, das Anlegen von größeren Dateien (z.B. Filemaker) zur Zitat-, Exzerpt-, Bildverwaltung. Mehr mache ich nicht.

N-2.2. Favorisieren Sie bestimmten Metadatenformate/-schemata oder Normdaten für die Erschließung?

- Können Sie dies kurz begründen?

Nein, ich bevorzuge keine, da ich mich viel zu wenig damit beschäftige. Ich möchte, aus rein quellenkundlicher und quellenkritischer Natur, möglichst komplette Metadaten über ein Digitalisat besitzen, je umfangreicher, desto besser.

N-2.3. Welche Informationen erwarten Sie von einer Dokumentation zu dem Digitalisat? (z.B. Provenienz, Informationen zum Digitalisierungsprozess, Informationen zum Digitalisierungsprojekt, ...)

Möglichst viele. Auch wichtig sind mir die Informationen zum Digitalisierungsprojekt selbst. Ich sehe das auch bei zu beantragenden Digitalisierungsprojekten immer wieder als Problem, dass einerseits die Notwendigkeit eines Projektes wichtig ist, andererseits es immer aber auch eine Provenienz oder Genealogie des Argumentes braucht, warum ein Digitalisat/Projekt eine Rolle spielt. Diese Begründung möchte ich später auch in der Webpräsenz und der Verfügbarmachung der Daten lesen können. Ich möchte immer etwas über den historischen Hintergrund und das Zustandekommen der Daten erfahren. Je ausführlicher das ist, desto besser.

N-2.4. Welche Verarbeitungsstufe (Rohdaten, strukturiert, Volltext, ...) ist für Sie am sinnvollsten:

- bei Textdigitalisaten?

- bei Bilddigitalisaten?

Das ist von Thema zu Thema und Gegenstand zu Gegenstand unterschiedlich. Manchmal reichen mir Rohdaten, manchmal möchte ich in Volltexten ganz spezifische Punkte aufsuchen. Ich glaube, diese Frage könnte ich leichter beantworten, wenn ich mich

konkreter mit einem Programm oder Projekt beschäftigte, dann könnte ich dies spezifisch für die jeweiligen Schritte beantworten. Ich finde das generell zu beantworten schwierig.

N-2.5. Wie wichtig ist für Sie die Verlinkung von Objekten untereinander?

Hier kommt es auf die Dichte an. Wenn die Verlinkung zu dicht ist, lenkt sie mich ab. Ist an entscheidenden thematischen Knotenpunkten verlinkt, dann ist es sinnvoll. Deshalb ist es wichtig, dass Programme und Dateien von Leuten angelegt werden, die auch von der Sache etwas verstehen.

Wenn jedoch jede zweite Zeile eine Verlinkung bietet, will ich dies nicht wissen und breche das Programm dann oft ab, da ich das von mir gelesene dann auf zwei Ebenen wahrnehmen muss und das möchte ich nicht.

Zwischenfrage: Würden Sie den angebotenen Links folgen?

Ich würde den Links folgen, wenn sie an ausgewählten Stellen erfolgen. Ich folge ihnen nicht, wenn sie zu oft vorkommen oder, z.B. durch starke Farbigkeit oder Pulsieren, zu aufdringlich sind. Dies würde mich nur stören.

N-2.6. Wie sollten Gedächtniseinrichtungen mit Dubletten umgehen: Beide Exemplare digitalisieren, um mögliche Unterschiede sichtbar zu machen, oder auf der Basis eines (oder mehrerer) Exemplare ein beispielhaftes Modell erstellen?

Da antworte ich als Historiker sofort: Es gibt keine Dubletten. Das ist in naturkundlichen und geisteswissenschaftlichen Bereichen so. Daher lautet meine Antwort hier, beide Exemplare digitalisieren, um mögliche Unterschiede sichtbar zu machen. Aber auch das ist eine Ermessensfrage. Je nachdem, worum es geht. Bei Digitalisierungsprojekten, bei denen die Materialität eine große Rolle spielt, da würde ich immer sagen gibt es keine Dubletten. Bei Digitalisierungsprojekten, wo Materialität keine Rolle spielt und es rein um den thematischen Zusammenhang geht, da reicht es, wenn man ein Exemplar digitalisiert. Aber auch hier müssten wir darüber sprechen, was für ein Exemplar gemeint ist. Insofern ist auch dies schwierig zu beantworten. Es kommt natürlich auf den inhaltlichen Zusammenhang an. Trotzdem würde ich sagen, da Digitalisate nie das materielle Objekt selbst ersetzen können, genau weil es um die Metadaten von Materialität geht, die ein Digitalisat notwendigerweise abschneidet, im Zweifel immer beide Exemplare digitalisieren zu lassen.

N-2.7. Durch die Digitalisierung gehen einige Informationen wie z.B. die Materialität der Quelle in der Nutzung verloren. Können umfangreiche Metadaten hier einen gewissen Ausgleich schaffen?

Nein, dies ist nicht möglich. Je nachdem, wie man Ausgleich definiert. Natürlich können mir viele Metadaten im Vergleich zu sehr wenigen Metadaten helfen, mir einen Eindruck von der Quelle zu verschaffen. Sie können jedoch nie das Original ausgleichen, da der sensomotorische Apparat, den ich als Forscher besitze, hat immer noch mehr Möglichkeiten, als mir das Digitalisat bereitstellen kann. Ich kann eine Gramm-Angabe lesen, das Buch jedoch in der Hand zu halten und dessen Gewicht zu spüren und im Vergleich zu anderen Objekten auf neue Assoziationen und weitere Forschungsideen zu kommen, das ermöglicht mir auch ein Metadatensatz nicht.

N-3. Verfügbarkeit und Nachnutzung.

N-3.1. In wie fern beeinflusst die digitale Verfügbarkeit der Quelle Ihre Quellenauswahl?

Gar nicht. Ich gehe immer noch über die Quelle selbst. Wenn ich zur Ergänzung anfangen zu suchen, dann sicherlich. Ich bin jedoch nicht darauf angewiesen, dass es die digitale Verfügbarkeit gibt.

N-3.2. Vertrauen Sie Digitalisaten aus Gedächtniseinrichtungen als wissenschaftlichen Quellen?

- Wenn ja: Nur unter bestimmten Bedingungen?
- Wenn nein: Warum nicht? Was könnte eine Gedächtniseinrichtung hier verbessern?

Je umfangreicher der Metadatensatz, je mehr Zeit (und damit Geld) in die Metadaten gesteckt wurde, desto mehr vertraue ich dieser Quelle.

N-3.3. Wie sieht Ihr persönliches Forschungsdatenmanagement aus? Stützen Sie sich auf Angebote z.B. von virtuellen Forschungsumgebungen, nutzen Sie Cloud-Dienste oder arbeiten Sie lokal?

- Welche Dateien benötigen Sie hierfür?
- Gibt es Anforderungen aus der Fachcommunity oder von Förderern?

Diese Frage fasziniert mich sehr und ich würde gerne viel avancierter mit meinen Daten umgehen und mich auch mehr hineinarbeiten. Aus gesundheitlichen Gründen kann ich mich damit aber nicht auseinandersetzen. Dies spiegelt jedoch nicht wieder, was ich denke, dass wichtig wäre.

Zwischenfrage: Wenn man Forschungsdaten nicht nur als digitale Daten aufgreift, sondern auch analoge Daten mit einschließt. Können Gedächtniseinrichtungen ihre analog arbeitenden Nutzer dann auch unterstützen?

Ich arbeite analog sehr viel mit Exzerpten und dem Besitz von Büchern. Sehr viel arbeite ich mit einer Forschungstopographie in meinem Arbeitsraum. Zu bestimmten Themen mache ich mir Stapel (Bücher, Exzerpte, Notizen, Kopien, ...). Wenn ich das Gefühl habe, es ist ein gewisser Sättigungsgrad eingetreten, dann nehme ich diesen Stapel und beginne, diesen abzuarbeiten. Mein Arbeitsraum ist dazu unabdingbar. Im zweiten Schritt, sobald ich beginne, mich tiefer einzuarbeiten, arbeite ich auch mit digitalen Quellen. Der erste Schritt ist jedoch immer analog und dreidimensional. Vielleicht fehlt mir die Phantasie, ich weiß jedoch nicht, wie ich dabei unterstützt werden könnte.

Die optimale Unterstützung für mich ist die Zusammenarbeit mit wissenschaftlichen Hilfskräften. Wenn ich eine Spur verfolge, arbeite ich immer im Schneeballsystem. Dieses ist besonders befriedigend, wenn es eine gewisse Schnelligkeit hat. Ich bin im direkten Austausch mit den wissenschaftlichen Kräften, die mir Texte und Verweise holen, welche ich in meinem Stapel ablege. So wie dieser gesättigt ist und ich anfangen zu arbeiten, arbeite ich nur noch alleine. Da kann mich eine Gedächtnisinstitution nicht unterstützen.

N-3.4. Wie wichtig ist Ihnen eine Downloadfunktion für die angebotenen Dateien? Für einige Dateien besonders?

Sehr wichtig. JSTOR oder anderes sind echte Erleichterungen. Gerade in der Arbeitsweise des Schneeballsystems ist dies wichtig. Ich lese einen Text und eine Fußnote verweist auf einen Aufsatz, diesen will ich dann sofort haben. Nicht so sehr, um ihn sofort zu lesen, sondern um reinzuschauen, dessen Relevanz für mich zu überprüfen oder zu schauen, ob das Argument wie im Text angegeben dadurch unterstützt wird oder die Angabe vielleicht etwas krude ist. Dies möchte ich sofort nachprüfen können und daher arbeite ich sehr viel mit Downloadfunktionen.

N-3.5. Benutzen Sie wenn möglich Harvesting, um große Datenmengen abzurufen?

Nein, nie.

Zwischenfrage: Beauftragen Sie ihre Hilfskräfte damit?

Nein, auch nicht.

N-3.6. Wie beurteilen Sie Open Access für Digitalisate (sowohl bezüglich Zugänglichkeit wie Nachnutzung)?

Wenn man mal die politische Dimension des Begriffes beiseite lässt, sondern nur die Idee des offenen, unbeschränkten Zugangs zu Digitalisaten betrachtet, dann mag dies sicherlich hilfreich sein. Das Problem, was ich hierbei sehe ist, dass unter der Chiffre des Open Access schon zahlreiche Kapitalisierungsprojekte laufen, die ich ablehne. Die Utopie des Open Access sehe ich vielfach als gescheitert an. Ich finde es nach wie vor eine hinreißende Idee, doch ich glaube, es war eine Idee der 90er. Ich glaube eher, dass es im Moment hoch problematisch ist und bin daher vollkommen gespalten, habe mich bisher jedoch noch nicht eingehend damit beschäftigt.

Zwischenfrage: Konkret auf Gedächtniseinrichtungen bezogen, denken Sie, dass diese ihre digitalisieren Quellen frei und öffentlich zugänglich machen sollten.

Wenn sie finanziell unabhängig bleiben, dann ja.

Zwischenfrage: Auch zur Nachnutzung frei? Was halten sie von Bildagenturen von Gedächtniseinrichtungen, die diese oft aus öffentlichen Mitteln entstandenen Digitalisate vermarkten?

Ich finde eine finanzielle Entschädigung in dem Maße, dass es Bearbeitungskosten sind, vollkommen adäquat. Es geht mir darum, dass solche Daten dann nicht von größeren Unternehmen kapitalisiert werden. Solange es in der Eigenregie der einzelnen Gedächtnisinstitutionen bleibt ist es vollkommen angemessen. Sie sollten sich jedoch nicht in irgendeiner Form vertraglich an große Unternehmen wie Google binden, um bspw. nach zehn Jahren die Rechte abzugeben. Dies lehne ich ab und würde die Daten dann auch nicht nutzen, sollte ich es wissen.

N-4. Wünsche.

N-4.1. Würden Sie als Nutzer gerne in Entscheidungsprozesse von Gedächtniseinrichtungen zur Digitalisierung einbezogen werden?

Nein, ich persönlich als Nutzer nicht. Wenn es interessierte und umfassend informierte Nutzer gibt, die auch begründete Argumente finden können, außer die, die man vielleicht in der Zeitung oder einigen Büchern liest, dann vielleicht.

Ich jedoch kann die Fragen nicht adäquat beurteilen und finde es schwierig, wenn sich Institutionen quasi demokratisieren und Entscheidungen nach unten weiterleiten, die hier auf ein ungenügendes Maß an Informationssättigung stoßen. Dann bringt dies überhaupt nichts.

Ich möchte nicht mit einbezogen werden. Ich will, dass die Gedächtnisinstitution diese Entscheidungen trifft und im Sinne einer staatlichen Institution löst.

N-4.2. Was sind für Sie die Forschungsdaten der Geschichtswissenschaft?

- Sind auch Metadaten Forschungsdaten?

Ja, selbstverständlich sind auch Metadaten Forschungsdaten, gerade in meinem Fachbereich wäre ich ein schlechter Historiker, wenn ich dies ablehnen würde.

Alles, was dazu gehört, dass am Ende ein begründeter, nachvollziehbar argumentierter wissenschaftlicher Text steht, sind Forschungsdaten. Alles was dazu gehört, diesen Text zu schreiben und ihn in der, von der wissenschaftlichen Gemeinschaft akzeptierten, nachweisbaren Form zu veröffentlichen. Aber auch die Art und Weise, wie ich den Text schreibe, z.B. setze ich Fußnoten oder nicht, gehört für mich dazu. Das heißt, der gesamte Komplex, um den Text entstehen zu lassen aber auch seine Form, in der er dann gelesen wird, gehören für mich zu den Forschungsdaten. Also auch Entscheidungen im Forschungsprozess.

N-4.3. Wünschen Sie sich die Verwaltung von Forschungsdaten (Korpora, Textauswertung, ...) mit Bezug zu Digitalisaten durch die entsprechende Einrichtung?

- Wenn ja: Was spricht gegen ein disziplinäres Repositorium?

- Sehen Sie dies als Teil der Aufgaben von Gedächtniseinrichtungen?

- Würden Sie ihre Forschungsdaten selber in einem Repositorium ablegen, welches auch die Originale beinhaltet oder eher in fachspezifischen Repositorien?

Wenn ich durch meine Arbeit nachnutzbare Forschungsdaten wie z.B. Transkripte und Erschließung historische Tagebücher erstelle und die Bibliothek/Gedächtniseinrichtung dann mit der Frage auf mich zukommt, ob ich bereit wäre, die Daten online zu stellen, dann selbstverständlich.

Zwischenfrage: Gehört es zu den Aufgaben einer Gedächtniseinrichtung, Daten aktiv von den Nutzern einzusammeln?

Ich glaube ja. Es ist Aufgabe einer Institution zu wissen, wer in ihr forscht. Daher sind Fachreferenten und Aufsichten so wichtig. Finanzielle Kürzungen und damit Personaleinsparungen sind daher hochproblematisch, weil damit eine ganz wichtige und wesentliche Schnittstelle zwischen Nutzern und Büchern verloren geht. Dies betrifft nicht nur den Zugriff der Nutzer auf den Bestand, sondern auch, wie die Bibliothek ihren Bestand erweitern und erschließbarer machen kann. Man benötigt immer auch das persönliche Gespräch, um Daten, die man selber erarbeitet hat, zu verschenken und nichts anderes tun Forscher, wenn sie ihre Daten verfügbar machen. Das ist ein Aushandlungsprozess. Hätte ich solch ein Tagebuch transkribiert, so würde ich mir

ausbedingen, dass erst wenn mein Buch erschienen ist, auch diese Transkription online gestellt wird.

N-4.4. Abschluss: Können Sie kurz skizzieren, wie für Sie das ideale digitale Angebot einer Gedächtniseinrichtung aussehen würde?

Aus genannten Gründen fällt mir diese Skizze natürlich schwer, da ich viel zu wenig über den State of the Art bescheid weiß. Ich glaube, eine ganz wichtige Eigenschaft, die für mich ein digitales Angebot einer Gedächtniseinrichtung mitbringen muss ist, dass sie immer auch noch durch Personen vermittelt wird. Ich möchte nicht auf bspw. eine Bibliothek stoßen, in der ich niemanden um Hilfe bitten kann. Ich möchte nicht nur auf Grund einer Oberfläche eine Einrichtung besuchen, bei der ich keine Hilfe an die Seite gestellt bekomme.

Zwischenfrage: Also keine menschliche Hilfe?

Genau. Ich will auch keine Chats haben, da will ich gar nicht erst rein. Dafür ist mir meine Zeit viel zu schade, da ich mir das alles erst selbst suchen müsste. Nein, ich möchte gerne mit kompetentem Personal über das digitale Angebot sprechen können. Ich glaube, das ist mein wichtigster Wunsch für digitale Angebote in Gedächtniseinrichtungen, dass ich sie nie ohne Personal vermittelt bekommen möchte. Ich würde digitale Einrichtungen ablehnen, wenn sie zu einer Personalkürzung führen.

9.9.5. *Interview Nutzer 2.*

Nutzer 2

07.06.2019

Aus persönlichen Gründen möchte Nutzer 2 nicht, dass die während des Interviews gemachten Aussagen als Transkript veröffentlicht werden.

9.9.6. Interview Dr. Andreas Kohring.

Dr. Andreas Kohring - Humboldt-Universität zu Berlin, Institut für Geschichtswissenschaften

11.06.2019

Nf-1. System.

Nf-1.1. Ist Ihnen als Sucheinstieg die Suche oder das Browsing/Facettierung wichtiger?

Die Suche. Also wir reden hierbei immer über den Bestand von digitalen Materialien, die 99% des Tuns ausmachen. Klar gibt es noch andere Dinge, die nice to have sind, aber das ist eine andere Anwendung. Ich rede von der Brot-und-Butter-Anwendung. Da geht es nicht um Facettierung und Browsing, sondern um eine klar strukturierte Suchmaske, um sehr sehr schnell riesige Textkorpora zu durchsuchen. Eigentlich brauche ich keine Metadaten für die Suche, das liegt aber an der Struktur der Thesauri der griechischen und lateinischen Sprache. Wenn man das Papier erstmal verstanden hat, dann braucht man die Metadaten nicht mehr für die Suche. Ich suche Stellen oder Zusammenhänge und die finde ich am schnellsten, wenn ich einen Text mit einer Standard-Antik-Tastatur eingeben kann. Dann kriege ich die Ergebnisse in einer Trefferliste und da ich die Quellsprachen kann, bin ich auch nicht abhängig von irgendwelchen Übersetzungen.

Nf-2. Anforderungen.

Nf-2.1. Welche Metadaten eines Digitalisates sollten auf jeden Fall, welche idealerweise erfasst werden?

Die Metadaten, wie sie Irvine in den 70er Jahren angefangen hat liegen vor, mehr braucht man eigentlich nicht. Es geht um Textdatenbanken, hier braucht man den Autor und den Werktitel und die sind schon seit Jahrhunderten normiert. Dann braucht man noch die Kapiteleinteilungen, aber die sind auch schon im Teubner-Standard seit dem 19. Jh. normiert. Das liegt also alles vor und liegt den Datenbanken zugrunde. Das weiß ich, brauche es aber nicht, da das Interface mir eine bessere Recherche mit look and feel ermöglicht und eine Recherche an einem riesigen Regal erspart, was Zeit kosten würde.

Nf-2.2. Welche Grundanforderungen stellen Sie an Bilddigitalisate (z.B. Farbigkeit, Vergrößerungsfunktion, Farbraum, Qualität, OCR...)?

Grundsätzlich arbeitet die Antikenforschung zu überwiegendem Teil mit Texten. Es gibt die eine oder andere interessante Übersetzung und Projekte. Und wenn ich aus wissenschaftshistorischer Perspektive in andere Forschungsbereiche schaue, um die Entwicklung eines Textes über die Zeit zu betrachten, dann hätte ich natürlich gerne OCR. In einer gut gerasterten Grafikdatei, die mir vom look and feel die Produktion des frühen Buchdrucks nahe bringt, bringt es mir zwar von der Haptik viel, wenn ich sehe, wie das gemacht worden ist, doch vom forschenden Erkenntnisinteresse her interessiert es mich nicht, da ich ja eigentlich am Text und nicht seiner Aufbereitung interessiert bin. Wenn Bibliotheken wie die BSB da andere Wege gehen, verstehe ich das, aber die haben ja auch andere Nutzer im Auge. Insofern sind OCR, Vergrößerungen etc. zwar super,

aber wenn ich referenzierbar über Chain Recognition quasi den Inhalt bekomme und gleich durchsuchen könnte, wäre das wunderbar.

N-1. System.

N-1.1. Nutzen Sie eher Portale (DDB, Europeana, ...) oder direkt die digitalen Angebote von Gedächtniseinrichtungen?

- Können Sie dies kurz begründen?

Ich nutze Portale sowie direkte Angebote. Informationen zu Portalen bekomme ich über das Haus oder über Clio, wo ich auch Mitglied bin, und sehe so auch, was Kollegen in anderen Fachebereichen tun. Das interessiert mich wissenschaftlich, es ist aber nicht fachspezifisch. Für die Alte Geschichte sind diese Seiten schön, haben aber nicht eine so hohe Relevanz wie andere Informationsträger. Museumsseiten sind gut, wenn man Exkursionen vorbereitet, aber das ist eher ein Add-On und dann intensiv für die Vorbereitung, seltener für den akademischen Alltag.

Zwischenfrage: Würden Sie die fachspezifischen Angebote der Alten Geschichte auch als Portale ansehen?

Wenn ich mir spezielle Angebote anschau, z.B. die Linksammlung von Udo Hartmann, welche von Institut für Geschichtswissenschaften der Humboldt-Universität gehostet wird, so sind dies super Angebote, die einem eine einfache Google-Suche, also den studentischen Weg, ersparen. Ebenso erspart man sich Wikipedia, wobei die deutschsprachige Wikipedia im Bereich Altertum durch den hauptamtlichen Redakteur Marcus Cyron, den ich vor vielen Jahren auch ausgebildet habe, sehr gut ist. Normalerweise zieht man ja immer die englische Wikipedia vor, doch im Bereich Altertum würde ich das nie tun. Durch die Hauptberuflichkeit und die vielen Interventionen ist hier einfach das Niveau drin, das manche Angebote von Brill in den Schatten stellt.

Aber ansonsten sind die klassischen Portale eher ablenkend. Meistens suche ich ja etwas und dadurch, dass ich schon so lange in der Forschung tätig bin, interessieren mich die Beifunde nicht. Ich habe eine relativ genaue Vorstellung, was ich suche und habe mir über die Jahre dazu Bookmarks gemacht, so dass ich persönlich viel schneller arbeiten kann.

Ich weiß aber, dass es für Lernende leichter ist, je strukturierter die Angebote sind. Das ist ein ganz anderer Ansatz. Z.B. die Angebote zu 2000 Jahre Varusschlacht 2009, hier wurde auch digital sehr viel publiziert, auch digitale Angebote etc. Das ist super, wenn man dazu einen Kurs gibt, so hat man eine Lernumgebung, die andere Leute hosten. Schön, auch für die Studenten, dass es das gibt.

N-1.2. Welche Funktionen in digitalen Angeboten von Gedächtniseinrichtungen nutzen Sie am meisten bzw. sind am wichtigsten für Sie?

Texte. Manchmal gibt es Dinge wie Inschriften, die in Kooperationen zwischen Epigraphikern und Informatikern aufbereitet wird, mit Umschrift, mehreren Layern, die das ganze gut visualisieren und sich gut in den Unterricht einbauen lassen. Das ist nett, aber eher die Ausnahme, meistens sucht man Textangebote, die klassischen lexikalischen

Ressourcen wie den Neuen Pauly oder Brill. Und wie gesagt, was man wirklich braucht, hat man einfach als Bookmark und durchsucht nicht jedes mal den gesamten Bestand, was bestimmt spannend wäre, aber einfach Nerven und Zeit kostet.

N-2. Anforderungen.

N-2.1. Mit was für digitalen Daten arbeiten Sie hauptsächlich?

Welch Überraschung: Textdaten.

Zwischenfrage: Und welche Formate dabei?

Zum Downloaden hab ich durchsuchbares PDF am liebsten, ganz klar. Mir reichen aber auch Online-Angebote, je nachdem, wie es gemacht ist. Zum Lesen und Aufnehmen reicht das. Ich suche ja nicht Volltexte bestimmter Werke, sondern in der Regel Zusammenhänge begriffsgeschichtlicher Art usw. Und Volltexte zu durchsuchen ist extrem umfangreich und aufwändig.

Aber wie gesagt, ich favorisiere die Formate, die am schnellsten darstellbar sind, die kürzeste Retrieval-Zeit haben und möglichst umfangreich sind. Schön ist auch eine gewisse Fehlertoleranz im Griechischen. Je nach Unter- und Oberlängen, Akzenten, etc. gibt es hier Unterschiede. In den Anfängen in den 80er Jahren gab es hier nur den exakten Treffer oder eben nicht. Inzwischen werden solche Aspekte aus der Suche herausgelassen und man kann als Nutzer selber entscheiden, was richtig ist.

N-2.2. Favorisieren Sie bestimmten Metadatenformate/-schemata oder Normdaten für die Erschließung?

- Können Sie dies kurz begründen?

Nein, da ich die Metadaten eigentlich nicht nutze. Ich könnte nachschauen und wüsste, wie ich damit umgehe, aber die Provider der Informationen haben einen Trust Value und sind in der Wissenschaft so etabliert, dass ich deren Arbeitsweise nicht nachprüfen muss. Ich spare hier Zeit.

N-2.3. Welche Informationen erwarten Sie von einer Dokumentation zu dem Digitalisat? (z.B. Provenienz, Informationen zum Digitalisierungsprozess, Informationen zum Digitalisierungsprojekt, ...)

Eigentlich keine. Ich lese mir immer durch, was es zum Digitalisierungsprozess gegeben hat, wie die Finanzierung gewesen ist, wer dahinter gestanden hat. Einfach, weil es mich von der Wissenschaftslandschaft her interessiert. Wenn ich die Verantwortlichen kenne, dann weiß ich natürlich auch über die Projektgeschichte bescheid. Aber wenn ich die Angebote als Screenshot in einer Vorlesung einbinden möchte, dann werde ich mir nicht die Dokumentation durchlesen, sondern das schaue ich mir bei anderer Gelegenheit an. Es ist dennoch spannend und schön solche Sachen zu wissen, weil es immer auch zeigt, warum das Produkt so ist wie es ist.

N-2.4. Welche Verarbeitungsstufe (Rohdaten, strukturiert, Volltext, ...) ist für Sie am sinnvollsten:

- bei Textdigitalisaten?

- bei Bilddigitalisaten?

Ich hab gerne Volltext, gerne auch strukturiert. Zwei Zeilen einer griechischen Quelle sind schön, aber ideal ist es, wenn unten eine Zeile mitläuft, die mir gleich die Zitation der Quelle, also z.B. Zeile 1, Kapitel 1 von Buch X, angibt. Das ist mehr als Plain Text und mehr als ein PDF-Scan.

Bei Bilddigitalisaten hab ich wenig Erfahrungen. Manchmal schaue ich mir die Roman Imperial Coinage Daten an. Die sind super, man kann zoomen, erhält die Informationen aus den Büchern direkt neben dem Bild und kann sich direkt eine Münzserie, mit oder ohne Metadaten, ausgeben lassen, um zu schauen, wie sich eine bestimmte Münze über die Zeit entwickelt hat.

N-2.5. Wie wichtig ist für Sie die Verlinkung von Objekten untereinander?

Am liebsten hätte ich eine Vollverlinkung, dann wäre ich nämlich überflüssig. Also eine Datenbank, die die gesamten Texte der Antike, unabhängig von ihrer Sprache, hat und dazu eine Verlinkung der Realenzyklopädie des 19. Jahrhunderts, die jedoch durch den Rechteinhaber Brill nicht digitalisiert wird. Dann könnte man die Textdatenbanken und die Enzyklopädien über cross references miteinander spielen lassen. Das Perseus Projekt der Tufts University ist mal in die Richtung gegangen. Hier wurden alle Objekte mit freien Rechten zusammengebracht. Dazu wurde eine Geschichte der Antike gegeben und von da aus Hyperlinks auf die daran geknüpften Repositorien gelegt. Das zeigt schon eine Richtung, in die die Entwicklung gehen könnte.

N-2.6. Wie sollten Gedächtniseinrichtungen mit Dubletten umgehen: Beide Exemplare digitalisieren, um mögliche Unterschiede sichtbar zu machen, oder auf der Basis eines (oder mehrerer) Exemplare ein beispielhaftes Modell erstellen?

In der Alten Geschichte gibt es keine Dubletten. Es gibt diese Dubletten in so fern nicht, da jedes Objekt einen eigenen Erkenntniswert hat. Auch wenn es z.B. Abschriften eines Volksbeschlusses sind, wovon einer in Spanien und einer in Ankara steht. Dies sind aber keine Dubletten, sondern Artefakte, die mir eine reichsweite Verteilung beweisen. Meine Lieblingsdatenbank ist z.B. eine epigraphische Datenbank und hier suche ich am liebsten X. Das ist der Legionsstempel der zehnten Legion und das gibt die meisten Treffer. Diese Legion hat eine Geschichte von 280 Jahren und hat in dieser Zeit im gesamten Imperium gedient. Der Stempel an sich, z.B. an Dachziegeln, sagt als Inschrift so nichts, aber wenn man genug X-Stempel findet, dann weiß man, wie die Dislokation der Truppen funktioniert. Und wenn man sonst keine Quellen hat, dann ist man dankbar darüber, diese Legion so verorten zu können. Und auch wenn es tausende solcher Dachziegel gibt, so sind es keine Dubletten, da sie je nach hermeneutischem System eine Botschaft tragen und bestimmte Annahmen belegen.

N-2.7. Durch die Digitalisierung gehen einige Informationen wie z.B. die Materialität der Quelle in der Nutzung verloren. Können umfangreiche Metadaten hier einen gewissen Ausgleich schaffen?

Ja, die Informationen gehen verloren und Metadaten können keinen Ausgleich schaffen. In situ ist in situ ist in situ. Das ist in der Altertumskunde so. Es ist immer etwas anderes,

ein Objekt direkt vor sich zu haben und berühren zu können, egal wie hoch die digitale Qualität ist. Auch eine digitale Rekonstruktion des Forum Romanum kann nie die Eindrücke ersetzen, die man während einer Exkursion dorthin erhält.

Mag sein, dass die Technik es irgendwann ermöglicht, auch andere Eindrücke wie Riechen, Schmecken, Fühlen zu vermitteln, aber im Moment ist die Autopsie immer noch ein nicht ersetzbarer Wert an sich.

N-3. Verfügbarkeit und Nachnutzung.

N-3.1. In wie fern beeinflusst die digitale Verfügbarkeit der Quelle Ihre Quellenauswahl?

Darf sie eigentlich nicht, dass weiß jeder, der mal ein Tutorium besucht hat. Aber in der Praxis wird natürlich alles bevorzugt, wofür ich nicht aufstehen, nicht extra irgendwo hin fahren muss, sondern was einfach da ist. Das ist einfach eine rationale Entscheidung. Die reine Wissenschaftslehre sagt zwar was anderes, aber in der Praxis ist es etwas anderes.

N-3.2. Vertrauen Sie Digitalisaten aus Gedächtniseinrichtungen als wissenschaftlichen Quellen?

- Wenn ja: Nur unter bestimmten Bedingungen?
- Wenn nein: Warum nicht? Was könnte eine Gedächtniseinrichtung hier verbessern?

Ja, tue ich. Nicht allen, das hängt immer davon ab, aber die Gedächtniseinrichtungen, mit denen wir zu tun haben, da ist das mittlerweile anerkannt. Das war immer ein Problem, aber das hat man bei Printausgaben auch. Dann hat man konkurrierende Ausgaben und das Original-Manuskript ist nicht einsehbar. Hier muss man dann die Entscheidung treffen, ob man der Angabe vertraut und wenn man Pech hat und sie falsch ist, dann ist das Argument mal soeben pulverisiert worden. Das Problem selber ist ein Problem des Umganges mit Quellen und ihren Editionen. Die mediale Frage, digital oder analog, ist dabei nicht so erheblich.

Will eine Einrichtung die Akzeptanz in der Community verbessern, so geht dies natürlich über Zeit oder Personal sowie eine Etablierung in einem Verbund. Aber ansonsten kann ich dazu nichts sagen.

N-3.3. Wie sieht Ihr persönliches Forschungsdatenmanagement aus? Stützen Sie sich auf Angebote z.B. von virtuellen Forschungsumgebungen, nutzen Sie Cloud-Dienste oder arbeiten Sie lokal?

- Welche Dateien benötigen Sie hierfür?
- Gibt es Anforderungen aus der Fachcommunity oder von Förderern?

Ich nutze zwar auch Cloud-Dienste, aber lieber arbeite ich lokal. Warum? Weil die Plattenspeicher nichts mehr kosten und immer verfügbar sind. Ich kann eine handgroße ein Terabyte SSD mit mir herumtragen und kann diese an praktisch jeden PC anschließen und brauche so das Netz nicht, muss mir nicht die Fragen stellen, ob das Netz funktioniert, ich mich bei Eduroam anmelden kann usw. sondern habe die Daten dabei. Aber das liegt glaube ich an meinem Alter. Es gibt viele jüngere Kollegen, die arbeiten

komplett cloud based. Das heißt nicht, dass ich Clouds nicht nutze, aber wenn ich die Wahl habe sind mir lokale Speicher lieber. Aber das ist Geschmackssache.

N-3.4. Wie wichtig ist Ihnen eine Downloadfunktion für die angebotenen Dateien? Für einige Dateien besonders?

Eine Downloadfunktion ist schön, aber ich kenne die Rechte und weiß, dass es mal möglich ist und mal eben nicht.

N-3.5. Benutzen Sie wenn möglich Harvesting, um große Datenmengen abzurufen?

Nein, unsere Datenmengen sind zu klein. Harvesting heißt für mich, dass ich mit Massendaten arbeite. Meine Textdatenbanken enthalten zwar Massen an Daten, sind aber keine richtigen Massendaten, da die Fragen, die ich daran habe den Bestand schnell ausgeben. Die Ergebnisliste beträgt dann vielleicht $n=25$, da brauche ich kein Harvesting, da brauche ich noch nicht mal die Suche verfeinern, denn 25 Artefakte kann ich mir anschauen. Bei der Größe dauert das Schreiben einer neuen Query genauso lange, wie einfach einmal rüberzuschauen.

Das sieht bei Zeithistorikern völlig anders aus, aber ich arbeite hier mit endlichen Quellenbeständen. Und da ich weiß, dass es endlich ist, will ich auch alles einmal gesehen haben. Insofern brauche ich nicht große Maschinen, die mir Submengen oder Querschnittsmengen ausgeben, da fehlt hier einfach der Bestand in der Größenordnung.

N-3.6. Wie beurteilen Sie Open Access für Digitalisate (sowohl bezüglich Zugänglichkeit wie Nachnutzung)?

Die Open Access Idee ist eine super Idee. Und wie schon gesagt, die Idee eines Perseus Projektes, welches nicht bei Tufts College gehostet ist, sondern die Community umfasst und keine nationalen Grenzen, (Urheber-)Rechte und schützenswerte Verlagsinteressen mehr kennt, sondern sich alle weltweit auf eine Commons-Lizenz oder ähnliches geeinigt hätten ist phantastisch. Dann stellt sich aber immer die Frage, wenn dies alles ohne Zugangskosten vorhanden ist, wer garantiert die Validität der Daten, die nachhaltige Benutzbarkeit usw. Es gibt ja auch viele Probleme, die bei Open Access Ressourcen auftauchen, eben wenn die Leute dabei auf die Idee kommen zu sparen. Ich halte Open Access nicht für eine Methode zum Sparen, es ist eine andere Wissenschaftsidee. Ähnlich wie Wikipedia. Eigentlich steckt da Demokratie hinter, nicht Sparen, nicht Klauen oder Plagiate. Schwarmintelligenz und Demokratie gehen zusammen und bilden dabei das Ideal der Wissenschaft im Menschheitsauftrag. Das ist die Idee, die eigentlich dahinter steckt. Und in der besten aller Welten, in der wir nicht leben, aber die ich mir vorstellen kann, wäre das natürlich super. Da ist das alles von den Steuerzahlern irgendwann mal irgendwie bezahlt worden und sollte daher für alle zukünftigen Steuerzahler verfügbar und nicht durch verschiedene Verlagsrechte eingeschränkt sein.

N-4. Wünsche.

N-4.1. Würden Sie als Nutzer gerne in Entscheidungsprozesse von Gedächtniseinrichtungen zur Digitalisierung einbezogen werden?

Nein, muss ich nicht.

N-4.2. Was sind für Sie die Forschungsdaten der Geschichtswissenschaft?

- Sind auch Metadaten Forschungsdaten?

Natürlich sind Metadaten auch Forschungsdaten. Alles was ich mit der Forschung verknüpfe sind Forschungsdaten, ganz klar.

Zwischenfrage: Also z.B. auch Entscheidungen, die im Forschungsprozess getroffen werden?

Genau. Es ist natürlich immer die Frage. Bei Publikationen ist es völlig klar, es gilt das gedruckte Wort. Wenn man aber an neue Kooperationsmodelle, Sonderforschungsbereiche etc. denkt, dann ist eigentlich auch der Weg zu einem Sammelband oder was am Ende rauskommt Forschungsdaten und sollten vorgehalten werden und nachvollziehbar sein. Wie gesagt, wenn man eine Unternehmensgeschichte schreibt, dabei das Unternehmensarchiv digitalisiert und anschließend veröffentlicht, dann ist das perfekt. Man muss nur ein Unternehmen finden, das dies mitmacht, aber in der besten aller Welten wäre es super. Problematisch sind Unternehmensarchive, die nicht frei verfügbar sind. Die Quellen hier können zwar zitiert und referenziert werden, aber niemand in der Wissenschaft kann diese Quellen nachprüfen, ohne selber einen großen Aufwand betreiben zu müssen. Aber das sind nicht nur Urheberrechtsfragen, sondern auch Persönlichkeitsschutz, Firmeninteressen etc. Wie gesagt, in der besten aller Welten könnten diese Prozesse nach 70 Jahren veröffentlicht werden, aber die reale Rechtsprechung ist hier ganz anders.

Wie gesagt, ich bin der Meinung, Forschungsdaten sind mehr als die Publikation und alles was in der Forschung entsteht, sollte auch nachnutzbar sein und nicht einfach wie ein Datenfriedhof werden, der nach Projektende nicht mehr beachtet und anschließend weggeschmissen wird. Das ist meiner Meinung nach Blödsinn. Und demnach wären auch Metadaten Forschungsdaten, klar.

N-4.3. Wünschen Sie sich die Verwaltung von Forschungsdaten (Korpora, Textauswertung, ...) mit Bezug zu Digitalisaten durch die entsprechende Einrichtung?

- Wenn ja: Was spricht gegen ein disziplinäres Repositorium?

- Sehen Sie dies als Teil der Aufgaben von Gedächtniseinrichtungen?

- Würden Sie ihre Forschungsdaten selber in einem Repositorium ablegen, welches auch die Originale beinhaltet oder eher in fachspezifischen Repositorien?

Klar, mit Bezug zu den Digitalisaten durch die entsprechende Einrichtung wäre super. Ich weiß aber nicht, ob es Aufgabe der Einrichtung ist. Es muss eine Einrichtung geben, die das macht, aber ich weiß nicht, ob es die Gedächtniseinrichtung sein muss. Natürlich könnte man sich auch auf einen Weltstandard einigen und sagen, das wäre eine supranationale Aufgabe im Sinne der UNESCO und alle Einrichtungen mögen ihre Rechte daran abtreten. Das wäre schon gut, man bräuchte nicht mehr einzelne Einrichtungen und es ginge schneller.

Das ist aber eigentlich eine politische Aufgabe, nicht Aufgabe einer einzelnen Einrichtung. Und diese Aufgabe wäre wahrscheinlich auch nicht national, sondern eben wirklich auf einer weltweiten Ebene. Eine einzelne Einrichtung würde sich Regeln machen, die sich

wieder von den Regeln anderer Einrichtungen unterscheiden. Als Forscher habe ich dadurch nichts gewonnen.

Klar, ich würde meine Forschungsdaten auch bereitstellen, habe ich auch schon gemacht. Aber ich darf die Scans von Archivakten nicht bereitstellen, dazu fehlen mir die Rechte. Als Forscher darf man zwar auf die Akten zugreifen, sie aber eben nicht direkt veröffentlichen.

N-4.4. Abschluss: Können Sie kurz skizzieren, wie für Sie das ideale digitale Angebot einer Gedächtniseinrichtung aussehen würde?

Gut, die ideale, die beste aller Welten habe ich ja oben schon beschrieben.

9.10. Vorlage für Einverständniserklärungen der Interviewten.

Einverständniserklärung zur Masterarbeit „Quellen digitalisieren, Digitalisate als Quellen - Anforderungen an und Management von Retrodigitalisaten in Gedächtniseinrichtungen als Forschungsdaten der Geschichtswissenschaft“ von Herrn Jan Wierzoch.

Ich erkläre mich einverstanden, dass mein Name und die im Interview gemachten Aussagen als Transkript in der Masterarbeit „Quellen digitalisieren, Digitalisate als Quellen - Anforderungen an und Management von Retrodigitalisaten in Gedächtniseinrichtungen als Forschungsdaten der Geschichtswissenschaft“ genannt werden.

Ja ☐

Nein ☐

Im Falle einer Veröffentlichung der Arbeit erkläre ich mich mit der Nennung meines Namens einverstanden.

Ja ☐

Nein ☐

Im Falle einer Veröffentlichung der Arbeit erkläre ich mich mit der Veröffentlichung des Interview-Transkriptes einverstanden.

Ja ☐

Nein ☐

Möchten Sie nach Abschluss der Arbeit ein Beleg-/Ansichtsexemplar erhalten?

physisch ☐

digital ☐

Nein ☐

Berlin, den